



Facultad de Ingeniería
UNIVERSIDAD NACIONAL de LA PLATA

**DETECCIÓN DE FALLAS EN SISTEMAS
DE CONTROL AUTOMÁTICO.
APLICACIÓN A UNA PLANTA
DE TRATAMIENTO DE EFLUENTES**

Mariana Alejandra Suárez

Tesis presentada para obtener el grado de
MAGISTER EN INGENIERÍA

Diciembre 2002

Departamento de Electrotecnia,
Laboratorio de Electrónica Industrial,
Control e Instrumentación (*LEICI*)

UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE INGENIERÍA

Departamento de Electrotecnia

Lugar de Trabajo:

Laboratorio de Electrónica Industrial,

Control e Instrumentación (LEICI)

Director:

Dr. Carlos Horacio Muravchik

Jurados de Tesis:

Dr. Dardo Marqués (Instituto Tecnológico de Buenos Aires)

Dr. Jorge Alberto Solsona (Universidad Nacional del Comahue)

Resumen

En esta tesis se estudia la detección y aislamiento de fallas en sistemas de control automático, y su aplicación a una planta de tratamiento de efluentes.

Derivamos el modelo general no lineal para una planta de remoción biológica de carbono; en ella, se controla la concentración de oxígeno disuelto en el agua a tratar mediante un controlador proporcional + integral. Desarrollamos modelos no lineales para las perturbaciones usuales y las fallas más comunes.

Teniendo en cuenta que el proceso que estudiamos responde a un modelo esencialmente no lineal, centramos nuestro análisis en el diseño de observadores no lineales para ser utilizados como generadores de residuos. En efecto, ésta es una de las técnicas más usuales para la detección de fallas.

Una propiedad fundamental para un generador de residuos es poseer alta sensibilidad a las fallas y mínima frente a perturbaciones externas, pues esto disminuye la frecuencia de falsas alarmas. En consecuencia, analizamos la robustez de varios observadores no lineales. Para los procesos que estudiamos, lograr que la salida sea completamente independiente de entradas desconocidas o perturbaciones, impone severas restricciones al modelo que sólo se verifican en casos especiales. Por ello, proponemos un nuevo camino que posibilite la atenuación de perturbaciones a un nivel prefijado. Deducimos en forma teórica las condiciones que debe satisfacer el observador para obtener convergencia semiglobal, en un compacto de espacio de estados y perturbaciones, y finalmente ilustramos los resultados obtenidos mediante simulaciones.

Los resultados muestran que nuestro observador de alta ganancia, rediseñado para atenuar perturbaciones, puede ser implementado con muy buena performance con el objetivo de detectar las fallas previstas en una planta de tratamiento de efluentes.

Índice General

1	Introducción	1
1.1	Organización de la tesis	4
1.2	Glosario	4
1.2.1	Detección y aislamiento de fallas	4
1.2.2	Sistemas de tratamiento de efluentes	7
2	Revisión del estado del arte	9
2.1	Detección y aislamiento de fallas	9
2.2	Observadores no lineales	11
2.3	Sistemas de tratamiento de efluentes	12
3	Detección y aislamiento de fallas	15
3.1	Principios básicos de la detección de fallas basada en un modelo	15
3.1.1	Estructura general para la detección y aislamiento de fallas	15
3.1.2	Estructura general para la generación de residuos. Detectabilidad de fallas	21
3.1.3	Conjuntos de residuos para aislamiento de fallas y aislabilidad	24
3.1.4	Técnicas para la generación de residuos	27
3.1.5	El problema de la generación de residuos robustos	36
3.2	Generación de residuos robustos utilizando Observadores de Entrada Desconocida (OED)	39
3.2.1	Teoría y diseño	39
3.3	Esquemas para el aislamiento y detección de fallas basados en OED robustos.	44
3.3.1	Esquemas robustos para la detección de fallas	44
3.3.2	Esquemas robustos para el aislamiento de fallas	45
3.4	Detección de fallas en sistemas no lineales	48

3.4.1	Observadores no lineales para generación de residuos .	49
3.5	Otros enfoques para el análisis de la detección de fallas	56
3.5.1	Lógica difusa en el diagnóstico de fallas	57
3.5.2	Redes neuronales en el diagnóstico de fallas en SNL .	57
4	Desacoplamiento o atenuación de perturbaciones en sistemas de detección de fallas	59
4.1	Desacoplamiento de perturbaciones en sistemas lineales . . .	60
4.1.1	Desacoplamiento de perturbaciones en el dominio de la frecuencia	60
4.2	Desacoplamiento de perturbaciones en sistemas no lineales . .	65
4.2.1	Desacoplamiento total	65
4.2.2	Atenuación de perturbaciones	66
5	Sistemas de tratamiento de efluentes	72
5.1	Los tratamientos biológicos. Generalidades.	72
5.2	Principales métodos de pretratamiento y tratamiento primario de aguas residuales	75
5.2.1	Cribado	76
5.2.2	Sedimentación	76
5.2.3	Flotación	77
5.2.4	Neutralización y homogeneización	77
5.3	Tratamiento secundario: el proceso de lodos activos	78
5.4	Modelo matemático de la planta	83
5.4.1	Balance de masa para el oxígeno en el biorreactor . .	84
5.4.2	Balance de masa para el nutriente carbono (sustrato)	87
5.4.3	Balance de masa para los microorganismos en el clarificador	87
5.4.4	Balance de masa para los microorganismos en el biorreactor	87
5.4.5	Ecuaciones cinéticas	88
5.4.6	Ecuaciones del modelo	88
5.4.7	Cambio de coordenadas. El modelo sin perturbaciones	91
5.5	El modelo con perturbaciones	93
5.6	Algunas fallas típicas en las plantas de tratamiento de efluentes	95
6	Detección de fallas en una planta de tratamiento de efluentes	98
6.1	Observadores no lineales para el modelo en estudio	98
6.1.1	Observador de orden reducido	99

6.1.2	Observador de estructura variable	104
6.1.3	Observador del tipo Filtro de Kalman Extendido (TF-KE)	109
6.2	Estudio de la robustez del observador TFKE. Atenuación de perturbaciones.	114
6.2.1	Modelo de error	115
6.2.2	Convergencia del observador TFKE con atenuación de perturbaciones	115
6.2.3	Resultados y simulaciones	119
6.3	Análisis comparativo de los observadores propuestos	124
6.4	Detección de fallas en la planta	125
6.4.1	El modelo con fallas	125
6.4.2	Evaluación del residuo	128
6.4.3	Simulaciones y resultados para el proceso con fallas	129
7	Conclusiones	135
7.1	Conclusiones	135
7.2	Breve síntesis	137
7.3	Trabajo futuro	139
A	Lema de Meyer-Kalman-Yacubovich	140
B	Lema generalizado de Gronwall	142
C	Teorema de Artstein-Sontag	143
D	Acerca de la Ecuación 6.27	144
E	Obtención de la cota para la norma de $E(w)$	147
E.1	Región de operación elegida para establecer la cota de $E(w)$	147
E.2	Figuras para la determinación de la cota	148
	Bibliografía	152

Capítulo 1

Introducción

Debido a la creciente complejidad de los sistemas de control modernos, aumenta día a día la necesidad de la supervisión y el diagnóstico de fallas para proveer a dichos sistemas de mayor confiabilidad y eficiencia. Distintos aspectos, tales como la seguridad en la operación, la eficiencia y la protección ambiental, adquieren importancia no sólo en sistemas críticos tales como reactores nucleares y aviones, sino también en otros procesos como por ejemplo plantas químicas, en donde las fallas pueden ocasionar pérdidas económicas o un fuerte impacto en el medio ambiente.

Por otra parte, indicaciones tempranas acerca de fallas incipientes en el sistema pueden evitar la parada de la planta y el aumento de costos operativos y de mantenimiento que conllevan estas situaciones irregulares.

Existen ciertos criterios para evaluar la eficiencia de las distintas técnicas de detección de fallas en función de elegir la más adecuada para resolver un problema en particular. Los más importantes son:

1. rapidez de la detección
2. sensibilidad frente a fallas incipientes
3. falsas alarmas por unidad de tiempo
4. fallas que no se detectan (omitidas)
5. identificación incorrecta de fallas

Un método utilizado tradicionalmente en detección de fallas (DEF) consiste en la utilización de componentes físicos redundantes que cumplan idénticas

funciones. Mediante el empleo de actuadores, sensores y componentes repetidos, distribuidos espacialmente en toda la planta, se garantiza una cierta protección en presencia de una falla localizada. Otras técnicas de desarrollo más reciente eliminan la necesidad de la redundancia física, basándose en el hecho de que señales de naturaleza diferente procedentes de distintos sensores pueden emplearse en un esquema de comparación, estableciendo entre ellas relaciones funcionales. Entre estos métodos, llamados en general de redundancia analítica, existe una técnica que consiste en el diseño de observadores que generan una señal denominada *residuo*, obtenida como la diferencia entre el valor medido de la salida y el valor estimado por el observador.

Hay que destacar que si bien las técnicas de redundancia analítica se basan en modelos matemáticos del sistema supervisado, en general no es posible disponer de un modelo perfectamente preciso y completo de un sistema físico. Es usual que los parámetros del mismo varíen con el tiempo de manera incierta, y la característica de las perturbaciones (entradas desconocidas) y el ruido de medición no pueden modelarse con exactitud. En consecuencia, siempre habrá una diferencia entre el proceso real y su modelo, aun en ausencia de fallas. Estas discrepancias provocan dificultades metodológicas en las aplicaciones de la detección de fallas, siendo la fuente de falsas alarmas y de alarmas omitidas que degradan el desempeño del sistema de detección y aislamiento. Para evitar estos inconvenientes éste debe hacerse robusto, o sea insensible y hasta invariante frente a incertidumbres de modelado. A veces, sin embargo, esta reducción de la sensibilidad frente a perturbaciones trae aparejada una disminución de sensibilidad a las fallas. En consecuencia, una formulación más precisa de este problema implica aumentar la robustez frente a perturbaciones e incertidumbres del modelo pero sin perder (o incluso aumentando) la sensibilidad frente a las fallas. Un esquema que cumpla con estos requerimientos se denomina *esquema robusto de detección de fallas*.

El desarrollo de métodos robustos para DEF ha sido un tópico clave en las investigaciones de los últimos diez años, y se han propuesto variadas técnicas para resolver este problema. Sin embargo, sus aplicaciones prácticas todavía se encuentran en desarrollo, en especial para los sistemas no lineales que, como analizaremos más adelante, presentan restricciones en cuanto a la observabilidad y al desacoplamiento total de perturbaciones.

Tradicionalmente, el problema de DEF en sistemas no lineales se ha tra-

tado en dos pasos. En primer lugar, se linealiza el modelo en un punto de operación, y luego se aplican técnicas robustas para generar residuos. Este método sólo funciona correctamente cuando la linealización no introduce grandes errores con respecto al modelo original no lineal y el sistema opera en un entorno relativamente pequeño del punto de operación elegido. Para tratar sistemas con no linealidades fuertes y amplio rango de operación, el problema de DEF debe atacarse utilizando técnicas no lineales.

Como caso de aplicación, estudiaremos aquí la DEF en una planta de tratamiento de efluentes. El objetivo de éstas es evitar que los desechos domésticos e industriales vertidos en ríos y lagos alteren su ecosistema, lo cual ocasiona mortandad de peces y desaparición de especies vegetales. Estas y otras transformaciones biológicas y bioquímicas se producen por la introducción de sustancias o formas de energía extrañas a la composición natural del ecosistema acuático. En particular los residuos orgánicos e inorgánicos pueden inhibir o estimular el desarrollo de ciertos grupos de microorganismos y macroorganismos, pero el medio perturbado en su equilibrio dinámico tiende a recomponer las condiciones originales, generando un proceso de *autodepuración*. Esta misma dinámica, con tendencia a estados de menor energía del sistema, es utilizada en los tratamientos biológicos de aguas residuales. En lo esencial éstos consisten en un proceso de descomposición acelerada, por el cual se ofrece a los microorganismos responsables de dicha descomposición elementos y condiciones favorables, que normalmente constituirían factores limitantes del proceso.

Para poder diseñar un sistema de DEF para estas plantas se requiere, en primer lugar, un modelo apropiado, asociado al aspecto del comportamiento del proceso que se desea analizar. Para la descripción de la dinámica del sistema, hay que tener en cuenta que la reacción biológica es producida por una población mixta de microorganismos sobre un sustrato complejo, por lo cual su cinética no obedece a una sencilla relación enzima-sustrato. Además, estas características determinan que no sea posible expresar estequiométricamente dicha reacción. Como es de esperar, el comportamiento obedece a un modelo esencialmente no lineal.

Esta tesis tiene como objetivo diseñar un sistema de detección de fallas basado en observadores no lineales para la planta de tratamiento de aguas residuales, de modo que las fallas detectadas sean distinguibles de las perturbaciones o entradas desconocidas que puedan aparecer en la misma. Dicho de otra forma, se busca generar un residuo robusto frente a perturbaciones y

de alta sensibilidad a las fallas que pueden aparecer en el proceso. Veremos que el problema no es trivial, ya que las características del modelo impiden el desacoplamiento total de perturbaciones. En consecuencia analizaremos la posibilidad de su atenuación, con el objeto de minimizar su influencia en el residuo posibilitando una detección de fallas confiable.

1.1 Organización de la tesis

La tesis está organizada como sigue: a continuación, se presentan algunas definiciones útiles para la comprensión del problema, tanto en el campo de la DEF como en el de tratamiento de efluentes. En el **Capítulo 2** se muestra un panorama del estado del arte en los temas tratados. En el **Capítulo 3**, se presenta la teoría de detección de fallas para sistemas lineales y no lineales, tal como ha sido desarrollada por distintos investigadores. En el **Capítulo 4** se analiza brevemente el desacoplamiento y atenuación de perturbaciones, en especial en sistemas no lineales, estudiado para controladores por Isidori [33]. Luego se propone un nuevo problema, equivalente a lo que en sistemas lineales sería el “dual” del anterior: la atenuación de perturbaciones en observadores no lineales. Se extiende el análisis al caso, no contemplado en [33], en que las perturbaciones no son escalares. En el **Capítulo 5** se describen los sistemas de tratamiento de efluentes y se desarrolla el modelo matemático de la planta a utilizar. En el **Capítulo 6** se diseñan tres observadores no lineales para el proceso, y se presentan los resultados de la aplicación de la DEF a la planta, con las simulaciones correspondientes. Finalmente, en el **Capítulo 7** se muestran las conclusiones.

1.2 Glosario

En lo que sigue, definimos la terminología básica en las áreas de trabajo de esta tesis.

1.2.1 Detección y aislamiento de fallas

Falla [32]: Es una desviación no permitida de una propiedad característica que conduce a la incapacidad para lograr el objetivo propuesto. También puede definirse como un cambio inesperado de una función del

sistema, que puede no representar una avería física y que perturba la operación normal de un proceso automatizado. Una falla debe diagnosticarse tan pronto como sea posible incluso si es tolerable en su primera etapa, a fin de prevenir consecuencias serias.

Sistema de diagnóstico de fallas: Es un sistema de monitoreo utilizado para detectar fallas y diagnosticar su ubicación e importancia en el proceso. En general consiste en:

1. Detección de la falla (DEF): para realizar una decisión binaria *falla/no falla*.
2. Aislamiento de la falla: para determinar su ubicación (sensor, actuador, componente).
3. Identificación de la falla: para estimar su tamaño y tipo o naturaleza.

Redundancia física o de hardware: Las técnicas tradicionales de DEF se basan en métodos de redundancia física (o de hardware) que consisten en la utilización de múltiples sensores, actuadores, computadores y software para medir y/o controlar una variable particular. Sus mayores desventajas son la necesidad de equipamiento extra, los elevados costos de mantenimiento y el espacio adicional que se requiere para los equipos.

Redundancia analítica: Esta técnica utiliza relaciones analíticas (o funcionales) redundantes entre varias variables medidas del proceso monitoreado (ejemplo: entrada-salida, entrada-entrada, salida-salida). En un esquema como éste, no se introducen fallas adicionales de hardware, ya que no se requiere hardware complementario, por lo cual la redundancia analítica es potencialmente más confiable que la redundancia física.

Residuo: En los esquemas de redundancia analítica, la diferencia generada por las verificaciones de consistencia entre las diferentes variables se denomina señal residual. El residuo debería ser cero durante la operación normal, y distinto de cero cuando sucede una falla. De hecho esta propiedad del residuo se utiliza para determinar la ocurrencia o no de una falla.

Aislabilidad: El objetivo del aislamiento de una falla particular es distinguirla de otras determinando, como se dijo antes, su ubicación. Para

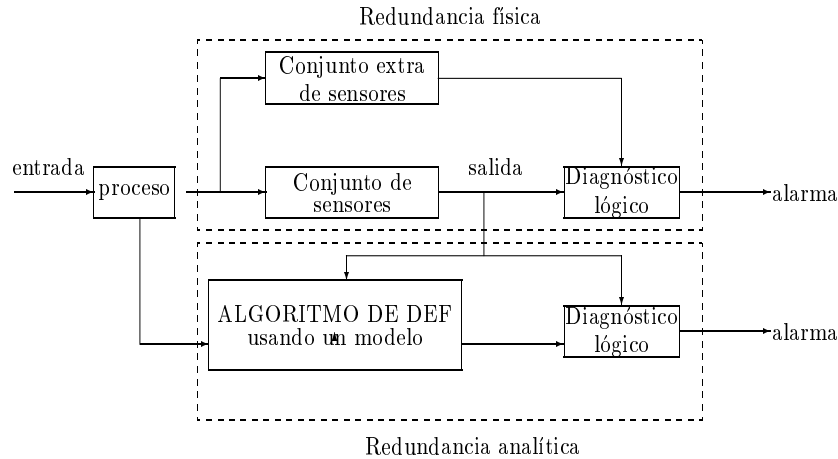


Figura 1.1: Redundancia física vs. redundancia analítica

esto, debe utilizarse un conjunto de residuos o vector residual. Una falla se dice *aislable* cuando puede distinguirse de otras utilizando un determinado vector residual. Si dicho vector puede aislar todas las fallas, posee la propiedad de *aislabilidad*.

Diagnóstico de fallas basado en un modelo [13]: El diagnóstico de fallas basado en un modelo puede definirse como la determinación de las fallas de un sistema por la comparación de mediciones disponibles del mismo con información previa representada por el modelo matemático del sistema, a través de la generación de señales residuales y su análisis.

Falla incipiente: Falla pequeña y de desarrollo lento, a veces llamada *falla suave*. Su efecto en el residuo no es muy notable, y puede permanecer oculta como consecuencia de incertidumbres del modelo. Una falla suave es una condición de mal funcionamiento que no es seria en un principio y frecuentemente se desarrolla en forma continua. No necesariamente produce una degradación significativa en el desempeño de la planta, pero su detección temprana puede evitar una avería seria en el sistema.

Estructura de un sistema de detección de fallas: Un sistema de DEF basado en un modelo comprende dos etapas principales [15]:

1. Generación de residuos: su objetivo es generar una señal residual que indique una falla, utilizando la información disponible de

entrada y salida del sistema monitoreado. En condiciones ideales el residuo es cero si no hay fallas, y notablemente distinto de cero cuando una falla sucede. En consecuencia, es independiente de las entradas y salidas del sistema. El algoritmo utilizado para producir esta señal se denomina *generador de residuos*. El residuo sólo debería aportar información de la falla, siendo idealmente invariante frente a perturbaciones. Además, para asegurar una DEF confiable, la omisión de fallas debería ser lo más pequeña posible.

2. Toma de decisión: Se examinan los residuos para determinar la probabilidad de fallas y se aplica una regla de decisión que permite definir si ha ocurrido una falla. El proceso de decisión puede consistir en un simple test de umbral para los valores instantáneos o promedios móviles de los residuos, o puede basarse en métodos de la teoría estadística de la decisión [64], [4].

1.2.2 Sistemas de tratamiento de efluentes

Calidad del agua: El agua contenida en cada río o lago se caracteriza por una cierta composición -mutable por ciertos efectos estacionales y otros no cíclicos- que definen la calidad del agua. Este no es un concepto absoluto, ya que cada uso y aplicación del agua exige propiedades particulares. Por ejemplo, la calidad del agua con miras a su utilización industrial no es la misma que si se piensa en la protección y reproducción de peces, ni tampoco se relaciona, por ejemplo, con su potabilidad. Estas diferentes calidades son definidas con parámetros específicos y las medidas de preservación de calidad tienden precisamente a mantener estos parámetros dentro de los valores aceptables en cada caso.

Demanda bioquímica de oxígeno (DBO): La respiración aeróbica resulta, necesariamente, en una demanda de oxígeno del ambiente. Por otra parte, la población de microorganismos en un ambiente dado es proporcional a la cantidad de alimento orgánico, biodegradable, en el mismo. Por ello, si se introduce una cierta cantidad de materia biodegradable en el agua, la demanda bioquímica de oxígeno es proporcional. El procedimiento formulado para su determinación (Royal Comission of Sewage Disposal, London, 1959), que se denominó *método de dilución*, ha sido utilizado, con pequeñas modificaciones, interna-

cionalmente como método estándar. Esencialmente, la DBO podría medirse colocando una muestra de agua en un frasco que se mantiene un cierto número de días a temperatura estándar. La DBO, en condiciones preestablecidas de tiempo y temperatura, es la diferencia entre la concentración inicial y la final de O_2 disuelto. Para poder realizar la determinación en casos de elevada carga orgánica, se diluye la muestra en diferentes proporciones de agua libre de materia reductora.

Demanda química de oxígeno (DQO): Se han ideado otros métodos para complementar o sustituir la DBO, considerando que la materia orgánica biodegradable del effluente es esencialmente reductora y por lo tanto, sujeta a oxidación. Estos métodos químicos consisten en oxidaciones controladas de modo de obtener una determinada correlación con los valores de la DBO. La medición de DQO permite una estimación en un tiempo menor y con métodos sencillos y de laboratorio.

Reactor biológico: El modelo de reactor que utilizaremos es el reactor de flujo continuo en tanque agitado, cuyas dos principales suposiciones son:

1. régimen estacionario
2. condiciones de mezcla completa

El lodo biológico se produce en el reactor en forma continua. Bajo las dos hipótesis anteriores, la concentración de DBO soluble en el mismo es igual a la del effluente (corriente de salida).

Floculación: El proceso por el cual la biomasa presente en el reactor forma masas granuladas que luego precipitan en la decantación secundaria es muy importante para la remoción de la materia orgánica. Las células de muchas bacterias poseen una vaina de consistencia gelatinosa constituida por polisacáridos, parcialmente soluble en agua, a la cual se ha atribuido la coagulación del material en suspensión para formar partículas mayores o flóculos. Sin embargo, algunos investigadores afirman que la coagulación se produce por las características coloidales de la masa de bacterias relacionadas con la intensidad de sus actividades metabólicas.

Capítulo 2

Revisión del estado del arte

Dado que se estudiará aquí la aplicación de la DEF a una planta de tratamiento de aguas residuales, revisaremos los trabajos más relevantes en las siguientes áreas vinculadas:

1. Detección y aislamiento de fallas
2. Observadores no lineales
3. Sistemas de tratamiento de efluentes y su control

2.1 Detección y aislamiento de fallas

Tanto la teoría como las aplicaciones de la DEF han sido temas de investigación de creciente importancia en los últimos veinte años. Este desarrollo se ha visto estimulado por la tendencia a una mayor automatización y complejidad de los procesos, así como la demanda de más confiabilidad y seguridad en los sistemas de control.

En los trabajos compilados por Basseville y Benveniste [5], varios autores desarrollan temas relacionados con la detección de cambios abruptos en señales y sistemas dinámicos. Por ejemplo, Willsky analiza la detección de cambios abruptos en procesos estocásticos mediante bancos de filtros de Kalman, mientras que Chow estudia las relaciones de redundancia en la detección robusta.

Un trabajo clave es el de Frank [25], que revisa el estado del arte en DEF hasta ese momento y presenta algunos nuevos resultados. Se enfatiza allí

en el uso de observadores de entrada desconocida para conseguir robustez, estudiando la posibilidad de desacoplamiento de perturbaciones. En el caso en que éste no sea viable, se plantean esquemas óptimos basados en índices de performance. Si bien este artículo es muy completo para el estudio de sistemas lineales, presenta sólo un esbozo del problema en sistemas no lineales.

La DEF en componentes y actuadores basada en observadores de entrada desconocida (OED) no lineales, expresando perturbaciones e incertidumbres de modelado como entradas desconocidas al sistema, se trata en Ge y Fang [29], y también en Seliger y Frank [52]. Si se satisfacen ciertas condiciones, el sistema se transforma mediante un cambio de variables en una forma que permanece inalterada frente a las perturbaciones, pero sí se ve afectada por las fallas. En otro trabajo, Li-Cheng Shen y Pau-Lo Hsu [38], estudian el diseño robusto de OED para DEF en los sistemas lineales.

Por su parte, Frank y Ding [23] presentan técnicas robustas para generación y evaluación de residuos, poniendo énfasis en las últimas contribuciones en el dominio de la frecuencia e inteligencia artificial. Alcorta García y Frank [1] muestran esquemas extendiendo las técnicas ya conocidas para sistemas lineales a los sistemas no lineales (SNL). Se discute desacoplamiento total de perturbaciones para ciertos SNL, así como la evaluación de residuos basada en la adecuada elección de una función umbral.

En Vemuri y Polycarpou [59], se describe un algoritmo de diagnóstico de fallas para una clase de SNL con incertidumbres de modelado donde no todos los estados del sistema son medibles, mientras que Yaz y Azemi [66] presentan un esquema para DEF para una clase de SNL continuos, usando ecuaciones y desigualdades lineales matriciales. La restricción para este método se encuentra, justamente, en la clase de SNL a la cual se aplica.

El trabajo de De Persis e Isidori [17] presenta una condición suficiente para la existencia de una transformación (difeomorfismo) que lleva al modelo a una forma en la que se deduce fácilmente la estructura que debe tener un generador de residuos para resolver el problema de detección de fallas en un sistema no lineal.

Algunos artículos presentan aplicaciones de la DEF a distintos procesos. En particular, Zolghadri y Monsion [68] desarrollan dos observadores no lineales para detectar fallas en una instalación hidráulica de laboratorio. La

técnica propuesta se restringe a cierta clase de sistemas no lineales. Aubrun, Harmand y Garnier [2], por su parte, proponen un esquema de DEF para un proceso de digestión anaeróbica, diseñando un observador por asignación de vectores y valores propios. El modelo utilizado es lineal, y se verifica la eficiencia del método de DEF en sensores y actuadores con datos experimentales de una planta piloto.

Otro enfoque del problema basado en la utilización de inteligencia artificial se presenta en el trabajo de Vega [58], donde se muestra un sistema de DEF mediante redes neuronales aplicado a una planta de tratamiento de efluentes real. Se diseñan dos experimentos y se toman datos, realizando la aplicación de los algoritmos fuera de línea.

2.2 Observadores no lineales

La teoría del control de sistemas no lineales y sus aplicaciones ha tenido un desarrollo creciente en los últimos años. En el libro de Isidori [33] se presenta la teoría de la realimentación no lineal, tanto para sistemas de una entrada y una salida como para aquéllos de múltiples entradas y salidas. Además, se desarrolla el enfoque geométrico para los sistemas no lineales, mostrando sus aplicaciones y estudiando la estabilización global con atenuación de perturbaciones en el caso SISO. Por otra parte en D'Attellis [16] se analiza la dinámica de los ceros para SNL y el diseño de observadores para sistemas no lineales con salida lineal.

Si bien en el último tiempo numerosos trabajos presentan técnicas para el diseño de observadores no lineales (ONL), mencionaremos aquí sólo algunos de ellos, en función de su relevancia para el estudio de nuestro sistema.

Walcott y Zak [60], por ejemplo, proponen observadores para SNL con incertidumbres o no linealidades acotadas, utilizando técnicas basadas en la teoría de los sistemas de estructura variable. En el artículo de Gauthier, Hammouri y Othman [28] se construye un ONL teniendo en cuenta ciertas hipótesis, como por ejemplo que algunas funciones sean globalmente Lipschitz. Se describe allí también una aplicación a sistemas biológicos. Por su parte, Mouyon [43] presenta una descripción de las técnicas de estimación de estados para SNL, haciendo énfasis en los observadores de alta ganancia y el filtro de Kalman. Schreier *et al.* [50] proponen un observador para una clase

de SNL. Además, un observador de alta ganancia para sistemas no lineales continuos y con ruido, con múltiples entradas y salidas, se desarrolla en el trabajo de García *et al.* [26].

Con respecto a las aplicaciones en sistemas biológicos, Farza, Busawon y Hammouri [22] desarrollan un observador para estimación de cinéticas en un biorreactor. Bosković [9] presenta un esquema de control adaptivo basado en un observador que requiere medición en línea de uno solo de los estados del proceso, aplicado a una clase de modelos no lineales que suele aparecer en sistemas con biorreactores. Finalmente, Biagiola *et al.* [8] proponen dos observadores diferentes para la estimación de la concentración de biomasa en bioprocesos discontinuos.

2.3 Sistemas de tratamiento de efluentes

Tal como en las secciones anteriores, presentaremos aquí una síntesis de los trabajos relacionados con el tema, teniendo en cuenta su aporte para el análisis de la detección de fallas en la planta de tratamiento de aguas residuales.

Los modelos que se presentan en la literatura para el tratamiento de aguas residuales se basan frecuentemente en aquéllos propuestos por la International Water Association (IWA). Esta organización realiza numerosas publicaciones, referidas a todos los aspectos del suministro y tratamiento de aguas, recolección y tratamiento de efluentes, y en especial a los estándares de cantidad y calidad de agua, incluyendo los requerimientos ambientales y de salud pública.

Una publicación reciente de la IWA [34] presenta un panorama de la familia de modelos propuestos para el tratamiento de lodos activados. Dichos modelos se denominan ASM1, ASM2 y ASM3. El primero de ellos es el que, hasta el momento, ha tenido mayor aplicación; describe la remoción de Carbono y Nitrógeno, y ha probado ser muy confiable. La dificultad para su utilización reside en la gran cantidad de parámetros que caracterizan al sistema, con lo cual las posibles formas de elección de los mismos son variadas. Para subsanar esta dificultad se han desarrollado modelos de orden reducido [36], en los cuales es relativamente más sencillo el estudio de propiedades tales como la controlabilidad y la observabilidad, así como

la identificación de parámetros. Por su parte, el ASM2 incluye mecanismos para la remoción de fósforo, es más complejo que el anterior y no ha sido tan extensamente probado como él. En cuanto al ASM3, el más reciente, en [34] se desarrolla su descripción y comparación con el ASM1, limitaciones y distintos aspectos de su aplicación.

Otros trabajos muestran algunas modificaciones a los modelos mencionados previamente. En [18], por ejemplo, se presenta un modelo para la descripción del comportamiento dinámico de la nitrificación y desnitrificación simultáneas en un reactor de lodos activados perfectamente mezclado, con aireación intermitente. En este caso, el ASM1 es extendido para incluir el proceso de volatilización de las impurezas del agua.

En cuanto al control de los procesos para la remoción biológica de nutrientes, un estudio detallado se encuentra en el libro de Olsson y Newell [44], en cuya primera parte (*Dinámica*) se trata el diseño y utilización de modelos para la planta. La segunda parte (*Diagnóstico*) desarrolla el problema del monitoreo y diagnóstico temprano de situaciones irregulares en el proceso, y la tercera (*Control*) muestra la implementación de estrategias de control para el mismo. La última parte se refiere a las mediciones en la planta.

Olsson y Aspegren [45] realizan una breve revisión histórica de los progresos en remoción biológica de nutrientes. Por su parte Carlsson *et al.* [11] muestran el diseño y operación de una planta piloto de lodos activados. En [31], además, se analizan diferentes métodos analíticos utilizados para la caracterización de efluentes y lodos.

En el artículo de Lindberg y Carlsson [39] un controlador adaptivo se aplica a una planta piloto para desnitrificación. En tanto, Carlsson y Milocco [10] presentan un controlador para la concentración de nitratos, cuando el objetivo es la remoción biológica de nitrógeno.

Una extensión de las leyes de control adaptivo en sistemas de una entrada y una salida a sistemas multivariados y su aplicación al proceso de lodos activados se muestran en el trabajo de Dochain [21].

Por su parte, Tenno y Uronen [56] consideran un proceso de tratamiento de efluentes donde tienen lugar simultáneamente la nitrificación y la remoción de carbono. Se calibra el modelo con datos de planta piloto y se desarro-

lla un algoritmo de control para controlar las concentraciones de amoníaco y materia orgánica.

En cuanto a la estimación de parámetros en el sistema, Lindberg y Carlsson [40] presentan un sensor para estimar la velocidad de respiración y la función no lineal de transferencia de oxígeno ($K_L a$). Además, la optimización de parámetros de diseño tales como la máxima velocidad específica de crecimiento microbiano (μ_{max}), la constante de saturación (K_S) y el coeficiente de decaimiento (k_d) en un sistema de lodos activados para el tratamiento de un effluente de la industria alimenticia, se muestra en el trabajo de Bertola *et al.* [7].

Un enfoque diferente del problema aparece en el artículo de Hong Zhao y Thomas Mc Avoy [67], quienes desarrollan un modelo híbrido para una planta de lodos activados para tratamiento de effluentes. Este integra un modelo simplificado de primer principio con un modelo de caja negra utilizando redes neuronales.

Capítulo 3

Detección y aislamiento de fallas

En este capítulo estudiaremos los principios básicos de la detección de fallas (DEF). Presentaremos algunas técnicas para la generación de residuos, destacando aquéllas que se basan en el uso de observadores. Dado que en los complejos sistemas de ingeniería que estudiamos es inevitable la presencia de perturbaciones o entradas desconocidas, adquiere importancia el análisis de la robustez de los residuos. Analizaremos entonces el diseño de los observadores de entrada desconocida (OED), cuyo principio es hacer que el vector de estimación de estados no se vea afectado por las perturbaciones.

Por último, estudiaremos la DEF en sistemas no lineales y la generación de residuos robustos mediante observadores no lineales (ONL), con las restricciones que se presentan en este caso para lograr el desacoplamiento total de las perturbaciones.

3.1 Principios básicos de la detección de fallas basada en un modelo

3.1.1 Estructura general para la detección y aislamiento de fallas

El diagnóstico de una falla basado en un modelo matemático puede definirse como la detección, aislamiento y caracterización de la falla a partir de la comparación entre mediciones disponibles del sistema y la información

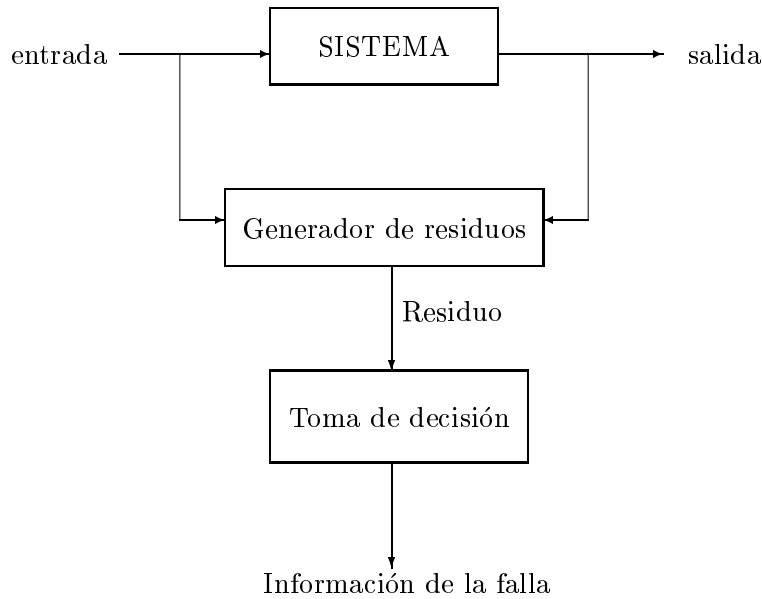


Figura 3.1: Estructura general para el diagnóstico de fallas

previa obtenida del modelo matemático del mismo.

Para detectar la falla, debe establecerse un umbral (fijo o variable) en el residuo generado por la diferencia entre las mediciones “reales” y sus valores estimados utilizando el modelo. El análisis posterior de cada residuo conduce al aislamiento de la falla.

La estructura conceptual de un sistema de diagnóstico de fallas basado en un modelo consta, tal como se ve en la Figura 3.1, de dos etapas principales: generación de residuos y toma de decisión [13].

El objetivo de la generación de residuos es producir una señal indicando la ocurrencia de una falla. El residuo debería ser nulo o cercano a cero en ausencia de fallas, pero diferente de cero cuando sucede una falla. En condiciones ideales, esto significa que el residuo no depende de las salidas y entradas del sistema, y sólo aporta información de la falla. Para asegurar una detección confiable, la pérdida de información de fallas en la generación de residuos debería ser lo más pequeña posible.

En cuanto a la toma de decisión, dicho proceso puede consistir en un

test sencillo utilizando valores instantáneos de los residuos comparados con un umbral, promedios variables de los residuos, u otros métodos estadísticos [5], [3].

Debido a incertidumbres en el modelo y a la presencia de perturbaciones el residuo no será, en general, nulo. El establecimiento de umbrales debe realizarse teniendo en cuenta el compromiso existente entre la elección de un umbral bajo (alta frecuencia de falsas alarmas) y uno elevado (disminución de la sensibilidad frente a las fallas).

Puesto que el diagnóstico de fallas se realiza durante la operación del sistema, requiere de mediciones en línea, es decir, la información de salida y entrada del sistema, disponible cuando éste se encuentra en funcionamiento. Dicha información consiste en la salida medida de los sensores y la entrada a los actuadores.

La salida medida se necesita generalmente en el control por realimentación, mientras que la entrada a los actuadores es la acción de control requerida, generada por el controlador y normalmente implementada en un sistema de cómputo apropiado.

En la Figura 3.2 puede verse que para la detección de fallas se utiliza el modelo a lazo abierto, con lo cual no es necesario considerar al controlador en el diseño de un esquema para DEF. Es decir que es posible diseñar en forma independiente el detector de fallas, dado que su dinámica sólo depende de la relación entrada-salida de la planta.

3.1.1.1 Modelado del sistema

El primer paso es construir el modelo matemático del sistema a ser monitoreado. Para ello utilizaremos el sistema a lazo abierto en el cual distinguimos, como puede verse en la Figura 3.2, actuadores, dinámica del sistema y sensores.

Por simplicidad en la descripción, trabajaremos en la primera parte del capítulo con sistemas lineales y extenderemos el análisis a los sistemas no lineales en la Sección 3.4.

En ausencia de fallas, la dinámica del sistema en variables de estado

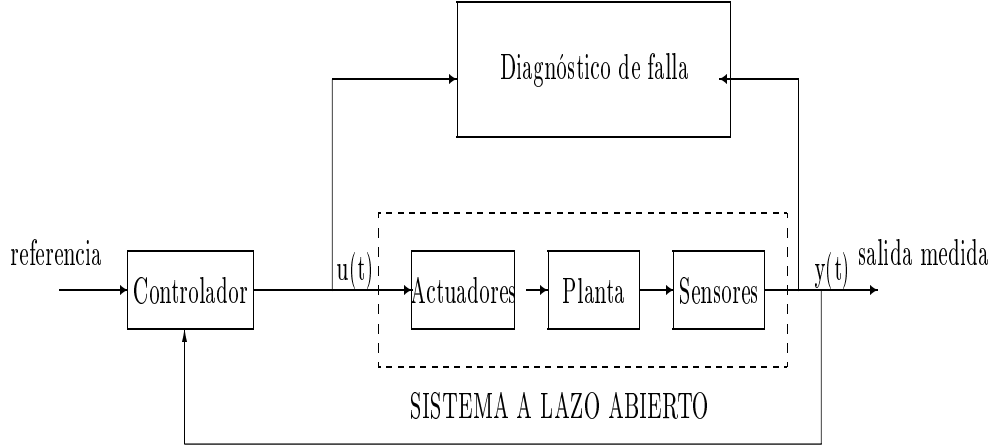


Figura 3.2: Diagnóstico de fallas y lazo de control

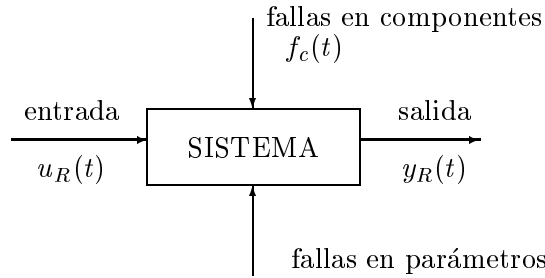


Figura 3.3: Dinámica del sistema

puede ser descripta por

$$\begin{aligned}\dot{x}(t) &= A x(t) + B u_R(t) \\ y_R(t) &= C x(t) + D u_R(t)\end{aligned}\tag{3.1}$$

donde $x \in \mathbb{R}^n$ es el vector de estados, $u_R \in \mathbb{R}^r$ es la entrada al actuador y $y_R \in \mathbb{R}^m$ es la salida del sistema (véase Figura 3.3). Por su parte A, B, C y D son matrices del sistema con dimensiones apropiadas.

Cuando hay una falla en los componentes, la dinámica del sistema puede describirse por

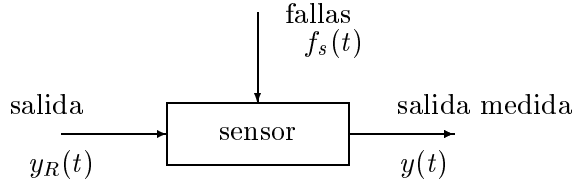


Figura 3.4: Fallas en los sensores

$$\dot{x}(t) = A x(t) + B u_R(t) + f_c(t) \quad (3.2)$$

La falla en un componente puede pensarse como un cambio en alguna condición del sistema, que modifica la ecuación dinámica. En algunos casos, podría expresarse como una variación de alguno de los parámetros del sistema, por ejemplo el elemento a_{ij} de la matriz A , y la ecuación dinámica quedaría

$$\dot{x}(t) = A x(t) + B u_R(t) + I_i \Delta a_{ij} x_j \quad (3.3)$$

donde

$$I_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

En general la salida del sistema y_R no es accesible en forma directa, por lo cual se utilizan sensores para medirla. Si se considera que la dinámica de dichos sensores puede despreciarse,

$$y(t) = y_R(t) + f_s(t)$$

siendo $f_s \in \mathbb{R}^m$ el vector de fallas del sensor (véase Figura 3.4). Eligiendo correctamente el vector f_s pueden describirse todas las situaciones que involucran fallas en los sensores.

Por otra parte, la señal $u_R(t)$ frecuentemente no es accesible. Para un sistema controlado, $u_R(t)$ es la respuesta del actuador a un comando $u(t)$,

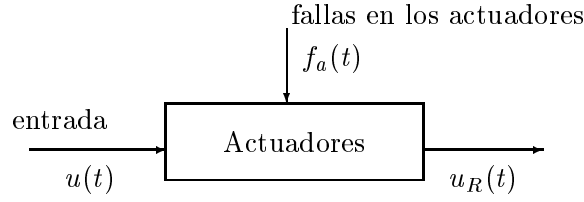


Figura 3.5: Fallas en los actuadores

que puede ser descripta (véase la Figura 3.5) por

$$u_R(t) = u(t) + f_a(t)$$

donde $f_a \in \mathbb{R}^r$ es el vector de falla del actuador y $u(t)$ es la señal de control conocida. Se asume en este caso que no se considera la dinámica de los actuadores, o que ésta queda absorbida en el modelo de la dinámica del sistema.

Del mismo modo que para las fallas en los sensores, diferentes situaciones de falla en los actuadores pueden representarse por una función apropiada de falla $f_a(t)$.

Para un sistema con posibles fallas en sensores, componentes y actuadores el modelo queda descrito por

$$\begin{aligned} \dot{x}(t) &= A x(t) + B u(t) + B f_a(t) + f_c(t) \\ y(t) &= C x(t) + D u(t) + D f_a(t) + f_s(t) \end{aligned} \quad (3.4)$$

Generalizando el modelo en el espacio de estados para todas las fallas posibles, puede expresarse

$$\begin{aligned} \dot{x}(t) &= A x(t) + B u(t) + R_1 f(t) \\ y(t) &= C x(t) + D u(t) + R_2 f(t) \end{aligned} \quad (3.5)$$

donde $f(t) \in \mathbb{R}^g$ es el vector de falla, correspondiendo cada elemento $f_i(t)$, ($i = 1, \dots, g$) a una falla específica. Desde un punto de vista práctico las fallas pueden considerarse como funciones desconocidas del tiempo. Las matrices R_1 y R_2 se conocen como matrices de entrada de las fallas, y representan el efecto de las mismas en el sistema. El vector $u(t)$ es la acción de control medida, y el vector $y(t)$ es la salida medida. Ambos son conocidos en el esquema de DEF.

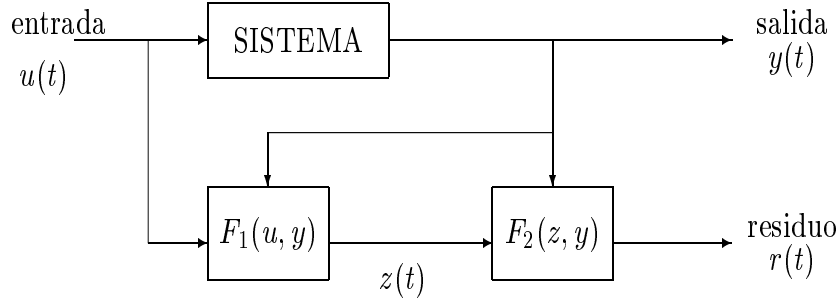


Figura 3.6: Estructura de señales redundantes

3.1.2 Estructura general para la generación de residuos. Detectabilidad de fallas

La contribución más significativa de los análisis modernos es la introducción de residuos que son independientes de los estados en los que opera el sistema y responden a las fallas de forma específica. Los residuos representan la inconsistencia entre las variables reales del sistema y el modelo matemático. En base a éste pueden establecerse relaciones invariantes (dinámicas o estáticas) entre distintas variables del sistema, y cualquier violación a dichas relaciones puede ser utilizada como residuo.

En términos de la estructura de señales redundantes la generación de residuos puede interpretarse [42], [4] así: el sistema $F_1(u, y)$ genera una señal auxiliar (redundante) z que junto con y genera el residuo r (véase la Figura 3.6) que satisface la siguiente relación invariante

$$r(t) = F_2(y(t), z(t)) = 0 \quad (3.6)$$

si no hay fallas.

Si hay fallas, esta relación invariante no se verifica y el residuo será no nulo.

El método más simple para la generación de residuos consiste en la duplicación del sistema, es decir, el sistema F_1 es idéntico al modelo original del sistema y tiene su misma señal de entrada. Es decir, el bloque F_1 es un simulador del sistema. Si bien este método presenta la ventaja de su simplicidad, su desventaja es que no se puede garantizar la estabilidad del

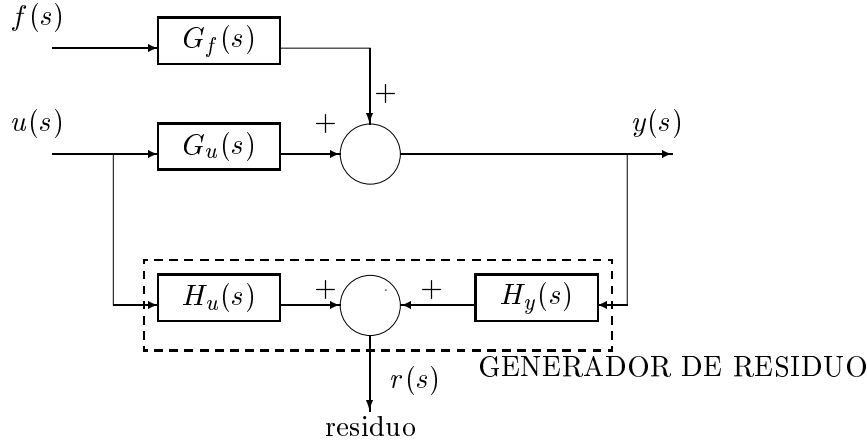


Figura 3.7: Estructura general de un generador de residuos

sistema F_1 si el sistema original es inestable, como consecuencia del uso del modelo a lazo abierto en el diagnóstico de fallas.

Una extensión del procedimiento anterior consiste en reemplazar el simulador por un estimador de la salida. En este caso, el sistema $F_1(u, y)$ utiliza ambas señales, u e y para generar una estima de una función lineal de la salida, por ejemplo, My . El sistema F_2 puede definirse como $F_2(u, y) = Q(z - My)$, donde Q es una matriz de peso (estática o dinámica).

Una estructura general se ve en la Figura 3.7, de la cual se obtiene

$$\begin{aligned} r(s) &= \begin{bmatrix} H_u(s) & H_y(s) \end{bmatrix} \begin{bmatrix} u(s) \\ y(s) \end{bmatrix} \\ &= H_u(s)u(s) + H_y(s)y(s) \end{aligned} \quad (3.7)$$

siendo H_u y H_y matrices de transferencia realizables utilizando sistemas lineales estables.

De acuerdo a la definición de residuo,

$$r(t) = 0 \quad \text{si} \quad f(t) = 0$$

Del diagrama de bloques de la Figura 3.7, puede verse que esta condición se verifica si

$$G_u(s) H_y(s) + H_u(s) = 0 \quad (3.8)$$

Entonces el diseño del generador de residuos consiste en la elección de las matrices H_y y H_u que satisfagan la ecuación (3.8).

Una falla puede detectarse comparando la función de evaluación del residuo $J(r(t))$ con una función umbral $T(t)$, con un test como el siguiente:

$$J(r(t)) \leq T(t) \text{ para } f(t) = 0$$

$$J(r(t)) > T(t) \text{ para } f(t) \neq 0$$

Si el test es positivo (la evaluación del residuo excede al umbral) puede suponerse que existe una falla. Hay distintas formas de elegir las funciones de evaluación y los umbrales. Por ejemplo, $J(r(t))$ puede ser la norma del vector de residuos y $T(t)$ una constante positiva (umbral fijo).

3.1.2.1 Detectabilidad de la falla

Analicemos ahora la respuesta del vector residual cuando se producen fallas.

$$r(s) = H_y(s) G_f(s) f(s) = G_{rf}(s) f(s) = \sum_{i=1}^g [G_{rf}(s)]_i f_i(s)$$

donde $G_{rf}(s) = H_y(s) G_f(s)$ es una matriz de transferencia de fallas que representa la relación entre residuo y fallas, $G_{rf}(s)(i)$ es su i -ésima columna y $f_i(s)$ es la i -ésima componente de $f(s)$.

Entonces, para que la i -ésima falla $f_i(s)$ sea detectable en el residuo r debe verificarse

$$[G_{rf}(s)]_i \neq 0$$

y ésta es la *condición de detectabilidad de la falla f_i en el residuo r* .

3.1.3 Conjuntos de residuos para aislamiento de fallas y aislabilidad

El procedimiento posterior a detectar la falla es su aislamiento, es decir, distinguir una falla particular de otras. Para ello se necesita un conjunto de residuos (o vector residual). Si dicho conjunto permite aislar todas las fallas, puede decirse que posee la propiedad de aislabilidad.

3.1.3.1 Conjunto de residuos estructurado

Un método para lograr el aislamiento de las fallas consiste en el diseño de un conjunto de residuos estructurado. Cada residuo se diseña para ser sensible a un subconjunto determinado de fallas e insensible a las restantes. El procedimiento consta de dos partes:

1. especificar las relaciones de sensibilidad e insensibilidad entre residuo y fallas, de acuerdo al objetivo de aislamiento.
2. diseñar el conjunto de generadores de residuo de acuerdo a las relaciones deseadas y especificadas previamente.

Si deben aislarse todas las fallas posibles, un conjunto de residuos puede diseñarse según las siguientes condiciones de sensibilidad a las fallas

$$r_i(t) = R(f_i(t)) \quad i \in \{1, 2, \dots, g\}$$

donde $R(\cdot)$ son relaciones funcionales.

Un umbral lógico sencillo puede ser usado para tomar una decisión acerca de la aparición de una falla específica, teniendo en cuenta que

$$r_i(t) > T_i \Rightarrow f_i(t) \neq 0 \quad i \in \{1, 2, \dots, g\}$$

donde T_i son los umbrales. Esta estructura es muy simple y permite detectar simultáneamente todas las fallas, pero en la práctica su diseño es complejo y frecuentemente imposible.

3.1.3.2 Conjunto de residuos generalizado

Un esquema más utilizado consiste en diseñar cada residuo para que sea sensible a todas las fallas excepto a una, es decir

$$\begin{cases} r_1(t) = R(f_2(t), \dots, f_g(t)) \\ \vdots \\ r_i(t) = R(f_i(t), \dots, f_{i-1}(t), f_{i+1}(t), \dots, f_g(t)) \\ \vdots \\ r_g(t) = R(f_1(t), \dots, f_{g-1}(t)) \end{cases}$$

Si todos estos residuos se generan mediante un banco de observadores, la estructura se conoce como “*esquema del observador generalizado*” [23]. El aislamiento se realiza, nuevamente, con un test de umbral como el que sigue

$$\begin{aligned} r_i &\leq T_i \\ r_j &\geq T_j \quad \forall j \in \{1, 2, \dots, i-1, i+1, \dots, g\} \end{aligned}$$

Entonces, $f_i(t) \neq 0$

3.1.3.3 Vector residual de dirección fija

Una forma alternativa para aislar fallas consiste en el diseño de un vector residual direccional, contenido en un subespacio del espacio residual y asociado a una falla específica. Es decir, hacer

$$r(t; f_i(t)) = \alpha_i(t) l_i \quad i \in \{1, 2, \dots, g\}$$

donde el vector constante l_i es la llamada dirección de firma de la falla i -ésima y α_i es una función escalar que depende del tamaño de la falla y su dinámica.

Con este esquema, el problema de aislar la falla se transforma en el de determinar cuál de las direcciones de firma de falla conocidas se encuentra más cerca del vector residual generado (ver Figura 3.8).

Para que el aislamiento sea confiable, la relación entre las fallas y su firma debe ser uno a uno.

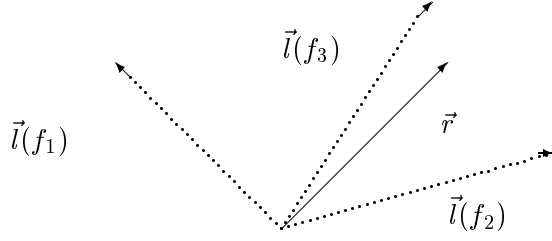


Figura 3.8: Vector residual direccional

3.1.3.4 Aislamiento de fallas en sensores y actuadores

Si sólo se consideran fallas en los sensores, la salida del sistema está dada por

$$y(s) = G_u(s) u(s) + f_s(s) \quad (3.9)$$

Para diseñar una señal residual sensible a un grupo de fallas $f_s^1(s)$ e insensible a otro $f_s^2(s)$, la ecuación anterior puede descomponerse como sigue.

$$\begin{bmatrix} y^1(s) \\ y^2(s) \end{bmatrix} = G_u(s) u(s) + \begin{bmatrix} f_s^1(s) \\ f_s^2(s) \end{bmatrix} \quad (3.10)$$

El generador de residuos tendrá la forma

$$r^1(s) = H_u^1(s) u(s) + H_y^1(s) y^1(s) \quad (3.11)$$

Sustituyendo $y^1(s)$ en (3.11), se obtiene

$$r^1(s) = (H_u^1(s) + H_y^1(s) G(s)) u(s) + H_y^1(s) f_s^1(s) \quad (3.12)$$

Para que el residuo sólo sea sensible al grupo de fallas $f_s^1(s)$, debe verificarse

$$H_u^1(s) = -H_y^1(s) G(s) \quad (3.13)$$

$$H_y^1(s) \neq 0 \quad (3.14)$$

Como puede verse, esta condición no es otra que la presentada en la ecuación (3.8) para un generador de residuos, es decir que el problema de aislamiento de fallas en sensores no plantea requerimientos adicionales.

La matriz $H_y^1(s)$ puede elegirse libremente en función de los requerimientos específicos, con la única restricción de que sea estable y realizable. Una

vez elegida, se determina $H_u^1(s)$ de la ecuación (3.13).

Como consecuencia de la libertad en la elección de $H_y^1(s)$, el aislamiento de fallas en sensores es siempre posible.

Cuando ocurren fallas en los actuadores, la salida del sistema queda

$$y(s) = G_u(s) (u(s) + f_a(s)) \quad (3.15)$$

Si se desea que el residuo sea sensible a un conjunto de fallas $f_a^1(s)$, e insensible a otro $f_a^2(s)$, la ecuación anterior puede descomponerse en

$$y(s) = G_u^1(s)(u^1(s) + f_a^1(s)) + G_u^2(s)(u^2(s) + f_a^2(s)) \quad (3.16)$$

El generador de residuos queda

$$r^1(s) = H_u^1(s) u^1(s) + H_y^1(s) y(s) \quad (3.17)$$

Sustituyendo $y(s)$ en (3.17) se obtiene

$$\begin{aligned} r^1(s) &= (H_u^1(s) + H_y^1(s) G_u^1(s)) u^1(s) + \\ &+ H_y^1(s) \{G_u^1(s) f_a^1(s) + G_u^2(s) (u^2(s) + f_a^2(s))\} \end{aligned}$$

Para que el residuo sólo sea sensible a $f_a^1(s)$ se necesita cumplir las siguientes condiciones.

$$H_u^1 = -H_y G_u^1 \quad (3.18)$$

$$H_y^1 G_u^2 = 0 \quad (3.19)$$

$$H_y^1 G_u^1 \neq 0 \quad (3.20)$$

De estas ecuaciones se infiere que para el problema de aislamiento de fallas en actuadores debe cumplirse una restricción adicional ($H_y^1 G_u^2 = 0$). No siempre existirá una matriz estable y realizable H_y^1 que satisfaga dicha restricción. Es decir, no hay libertad total para alcanzar el requerimiento de aislamiento de fallas en los actuadores y, en consecuencia, no siempre será posible dicho aislamiento.

3.1.4 Técnicas para la generación de residuos

La gran variedad de métodos existentes para la generación de residuos puede sintetizarse en [25]:

- a.- método del espacio de paridad
- b.- método del observador dedicado
- c.- método del filtro detector de fallas
- d.- método de identificación de parámetros

3.1.4.1 Método del espacio de paridad

La idea clave es verificar la consistencia de las ecuaciones matemáticas del sistema (relaciones de redundancia analítica) utilizando las mediciones. Se decide que ha ocurrido una falla cuando se sobrepasan los umbrales preasignados de error T_i .

Para delinear el concepto básico de la metodología del espacio de paridad, consideremos primero el caso simplificado de mediciones redundantes que pueden obtenerse directamente o de fuentes analíticas. Esto significa que se mide una magnitud por dos mecanismos diferentes, como por ejemplo, mediante instrumentos o sensores adicionales y de forma analítica (balances). Para el caso de una sola variable de salida (SISO), dichas mediciones se modelan con la ecuación algebraica de la medición

$$y = Cx + \Delta y \quad (3.21)$$

donde $y \in \mathbb{R}^q$ es el vector de medición, $C \in \mathbb{R}^{q \times n}$ es la matriz de medición de rango n , $x \in \mathbb{R}^n$ es el valor nominal de la medición y $\Delta y \in \mathbb{R}^q$ es el vector de error.

Si se cumple que $\Delta y_i > T_i$, se ha producido una falla en la variable i -ésima.

Para la detección de Δy el vector y puede ser combinado en un conjunto de ecuaciones linealmente independientes dado por

$$p = Vy$$

donde p es el vector dimensional de paridad y V es una matriz de proyección.

Con el objetivo de que p satisfaga el requerimiento usual para un residuo (es decir, ser nulo en ausencia de fallas), la matriz de proyección V se elige de modo que [25]

$$\begin{aligned} VC &= 0 \\ V^T V &= I_q - C(C^T C)^{-1} C^T \\ VV^T &= I_{q-n} \quad \text{con } q > n \end{aligned}$$

Esto significa que las filas de V son ortogonales y V es el espacio nulo de C . Entonces,

$$p = V \Delta y$$

Esta última ecuación revela que las ecuaciones de paridad contienen sólo los errores debidos a las fallas, independientemente de x que no es medida en forma directa. Además, en el espacio de paridad las columnas de V definen q direcciones diferentes de falla asociadas con cada medición. Por ejemplo, la i -ésima columna de V determina la dirección a lo largo de la cual cae p si

$$\Delta y = \Delta y_i = [0 \dots \Delta y_i \dots 0]^T$$

Esto asegura que una falla en la medición i implica un crecimiento de p en la i -ésima dirección. El vector residual q -dimensional $r = y - C \hat{x}$, donde $\hat{x} = (C^T C)^{-1} C^T y$ es la estimación de x por cuadrados mínimos y se relaciona con el vector de paridad p por

$$r = V^T p$$

Teniendo en cuenta todo esto, el problema de detección y aislamiento de fallas puede ser formulado como sigue: dadas q mediciones redundantes y_1, y_2, \dots, y_q de una variable de proceso, y cotas de error $T_1, \dots, T_i, \dots, T_q$ caracterizando el comportamiento de la falla,

1. Encontrar un estimado \hat{x} de las variables de proceso a partir de un subconjunto de mediciones.
2. Identificar la medición de la falla y chequeos de paridad. Obviamente, para detectar una falla única entre p componentes se necesitan $p - 1$ relaciones de paridad.

Generalizamos este concepto para el caso en que se usan relaciones de redundancia temporal de un sistema dinámico.

$$\begin{aligned} x(k+1) &= Ax(k) + Bu(k) \\ y(k) &= Cx(k) \end{aligned} \quad (3.22)$$

donde $x \in \mathbb{R}^n$ es el vector de estados, $u \in \mathbb{R}^p$ es el vector de entradas al actuador, $y \in \mathbb{R}^q$ es el vector de salida de los sensores, y A, B, C son matrices de dimensiones apropiadas.

Las relaciones de redundancia se especifican matemáticamente como sigue. Definamos al subespacio de vectores v de dimensión $q(s+1)$ por

$$P = \left\{ v / v^T \begin{bmatrix} C \\ CA \\ \vdots \\ CA^s \end{bmatrix} = 0 \right\} \quad (3.23)$$

que se llama *espacio de paridad de orden s* .

Cualquier vector v puede ser utilizado, en un instante k , para una verificación de paridad. El residuo es

$$r(k) = v^T \left[\begin{bmatrix} y(k-s) \\ \vdots \\ y(k) \end{bmatrix} - H \begin{bmatrix} u(k-s) \\ \vdots \\ u(k) \end{bmatrix} \right] \quad (3.24)$$

donde

$$H = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ CB & 0 & \dots & 0 & 0 \\ CAB & CB & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ CA^{s-1}B & CA^{s-2}B & \dots & CB & 0 \end{bmatrix}$$

Sustituyendo las ecuaciones de estado (3.22) en el residuo, se llega a

$$r(k) = v^T \begin{bmatrix} C \\ CA \\ \vdots \\ CA^s \end{bmatrix} x(k-s) \quad (3.25)$$

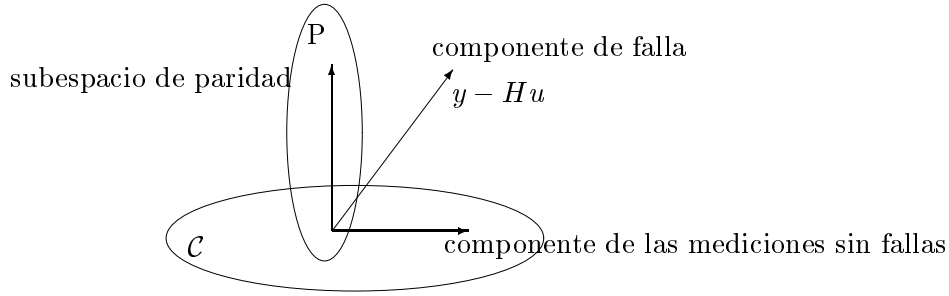


Figura 3.9: Interpretación geométrica del espacio de paridad

Dada la definición previa de v , ecuación (3.23), $r(k) = 0$ si no hay fallas.

De esto se concluye que una relación de redundancia es simplemente un modelo entrada-salida para una parte del sistema. En otras palabras, en lugar de verificar la consistencia del modelo matemático en su totalidad, nos restringiremos a comprobar relaciones individuales que son parte del modelo. Esto nos permite seleccionar las relaciones más confiables y proporcionar robustez al procedimiento de detección y aislamiento de fallas.

Para una interpretación geométrica de las verificaciones de paridad definidas precedentemente, consideremos el modelo (3.22), y sea \mathcal{C} el subespacio, ortogonal al de paridad P , generado por las n columnas de $[C \ CA \ \dots \ CA^s]^T$. Entonces, en ausencia de fallas, un vector $v = y - Hu$ se encuentra en el subespacio \mathcal{C} (Ver la figura 3.9).

Nótese que el método del espacio de paridad conduce a un tipo especial de observador para detección de fallas, llamado “dead-beat observer” [47].

3.1.4.2 Método del observador dedicado

La idea básica del método por observadores es reconstruir las salidas del sistema a partir de las mediciones o subconjuntos de ellas con la ayuda de observadores o filtros de Kalman utilizando el error de estimación o la innovación, respectivamente, como residuo para la detección y aislamiento de fallas.

A partir de la teoría de observadores, de gran desarrollo, se conoce que

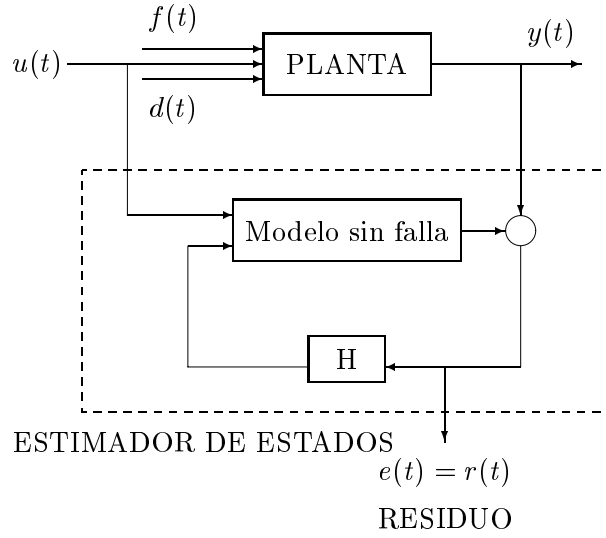


Figura 3.10: Configuración general para la estimación de estados.

la estimación de estados puede realizarse con observadores lineales y no lineales, de orden completo o reducido en el caso determinístico o filtros de Kalman en el caso estocástico, donde debe considerarse el ruido. La configuración fundamental para un observador lineal de orden completo se muestra en la Figura 3.10.

Notemos que el observador de orden completo consiste simplemente en un modelo paralelo del proceso con una realimentación del error de estimación, $e = y - \hat{y}$.

La realimentación es importante por las siguientes razones:

- para compensar diferencias en las condiciones iniciales.
- para estabilizar el modelo paralelo en el caso de un sistema inestable.
- para proveer grados de libertad en el diseño del filtro, por ejemplo, para desacoplar los efectos de algunas fallas de los de otras fallas o entradas desconocidas.

En el caso de un proceso lineal descrito por

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) + Ed(t) + Kf(t) \\ y(t) &= Cx(t) + Fd(t) + Gf(t)\end{aligned}\tag{3.26}$$

donde $d(t)$ son las perturbaciones y $f(t)$ las fallas, el estado \hat{x} y la salida \hat{y} de un observador de orden completo están dados por

$$\begin{aligned}\dot{\hat{x}}(t) &= (A - HC)\hat{x}(t) + Bu(t) + Hy(t) \\ \hat{y}(t) &= C\hat{x}(t)\end{aligned}\tag{3.27}$$

Donde el par (A, C) es observable y H es la matriz de ganancia del observador, elegida para obtener un desempeño deseado en el mismo.

Si se define el error de estimación de estados por $\xi(t) = x(t) - \hat{x}(t)$, teniendo en cuenta (3.26) y (3.27) las ecuaciones para su dinámica y la del error de salida del estimador, $e(t) = y(t) - \hat{y}(t)$, quedan dadas por

$$\begin{aligned}\dot{\xi}(t) &= (A - HC)\xi(t) + Ed(t) + Kf(t) - \\ &\quad - HFd(t) - HGf(t)\end{aligned}\tag{3.28}$$

$$e(t) = C\xi(t) + Fd(t) + Gf(t)\tag{3.29}$$

Se puede ver que el error de salida del estimador e es función de d y f , pero no de u . Entonces, puede ser usado como residuo con el objetivo de detectar y aislar fallas.

Cuando no hay fallas, $f = 0$, el residuo sólo es afectado por las perturbaciones, d . En cambio, si $f \neq 0$, el residuo varía, y se puede detectar una falla verificando el incremento del residuo debido a ella. En el caso más simple esto se puede hacer con un umbral adecuado. Para evitar falsas alarmas, el umbral debe elegirse lo más grande posible, aunque esto reduce la sensibilidad a las fallas. En forma similar se pueden generar residuos usando estimadores de orden reducido o no lineales. La idea fundamental en el diseño de estos estimadores para la DEF es su optimización basada en la elección correcta de la matriz de ganancia de realimentación H .

En lo que sigue, analizaremos distintas configuraciones posibles de los esquemas para DEF basados en observadores.

3.1.4.2.1 Estimador único (observador o filtro de Kalman) En este caso, un estimador único de orden completo o reducido se instrumenta con la salida del sensor más confiable y se reconstruye toda la salida. La comparación entre la salida medida y la estimada utilizando un umbral permite, en principio, la detección y el aislamiento de una falla de instrumento.

Para lograr mayor flexibilidad en el aislamiento de fallas de actuadores, componentes o sensores se puede diseñar un esquema de estimadores, o banco, cuya entrada sea el vector real de salida y o un subconjunto de él. Una técnica muy común se basa en testeo de hipótesis múltiples [64]. En este caso cada uno de los estimadores se diseña para una hipótesis diferente de falla. Las hipótesis se prueban en términos de funciones de verosimilitud usando, por ejemplo, teoría bayesiana de decisión.

Otra técnica conocida para la DEF consiste en asignar un estimador específico a cada uno de los sensores. En el denominado “Esquema del observador dedicado” [12], cada salida de sensor va a un estimador diferente y se estima el vector completo de salida o tantas componentes de él como sea posible.

3.1.4.2.2 Esquema del observador generalizado [25]. Esta es una versión alternativa que se diseña de modo que un observador destinado a determinado sensor tenga por entradas a todas las salidas menos la de dicho sensor. Este esquema permite detectar y aislar sólo una falla en uno de los sensores, pero con la ventaja de proveer mayor robustez frente a entradas desconocidas.

Por otra parte, puede combinarse redundancia analítica y de hardware con un sistema en “duplex” de sensores, en el cual se utilizan dos conjuntos idénticos de instrumentos, cada uno supervisado por un esquema de detección de alguna de las clases anteriores. Cuando ocurre una falla en un sensor, se detecta con la ayuda de los esquemas de observadores y el sistema se conmuta al sensor sano. La idea es que la DEF se realice mediante redundancia de hardware y el aislamiento con redundancia analítica.

En la Sección 3.4.1 analizaremos los esquemas para la detección y aislamiento de fallas utilizando estimadores no lineales.

3.1.4.3 Filtro detector de fallas

Inicialmente propuesto por Beard [6] y Jones [35], consiste en un estimador de estados de orden completo con una elección especial de la matriz H .

Sea un sistema con las siguientes ecuaciones de estado

$$\begin{aligned}\dot{x}(t) &= A x(t) + B u(t) + k_i f_i(t) \\ y(t) &= C x(t) + \bar{k}_j \bar{f}_j(t)\end{aligned}\tag{3.30}$$

donde $k_i \in \mathbb{R}^r$ es la dirección de diseño de la falla y r es la cantidad de direcciones de falla. Por su parte, f_i es una función arbitraria del tiempo, de tal modo que en ausencia de fallas, $f_i = 0$.

Las direcciones de falla k_i pueden ser utilizadas para modelar fallas en componentes y actuadores. En los sensores, las direcciones y modo de falla pueden modelarse con \bar{k}_j y \bar{f}_j .

Las ecuaciones del observador quedan

$$\begin{aligned}\dot{\hat{x}}(t) &= (A - HC) \hat{x}(t) + B u(t) + H y(t) \\ \hat{y}(t) &= C \hat{x}(t)\end{aligned}\tag{3.31}$$

La matriz de ganancia de realimentación H se elige de modo que el vector residual $r = y - \hat{y}$ tenga ciertas propiedades direccionales cuando ocurra determinada falla, es decir, se restrinja a una dirección única o plano en el espacio de residuos, independientemente del modo de f_i . Por lo tanto, la información importante para la detección está en la dirección del residuo más que en su variación temporal. Una falla se detecta cuando una o más de las proyecciones del residuo a lo largo de la dirección (o plano) de falla conocida es suficientemente grande.

Si bien el aspecto más atractivo del FDF es que la dirección de falla del residuo no se ve afectada por el tamaño o historia temporal (modo) de la falla, nótese que no se han considerado en el análisis las perturbaciones o entradas desconocidas, incluyendo variación de parámetros y ruido de medición. En consecuencia, este esquema requiere de un modelado preciso.

3.1.4.4 Identificación de parámetros

Este método se basa en el hecho de que las fallas de un sistema dinámico se reflejan en sus parámetros físicos (fricción, masa, viscosidad, etc.), lo cual permite detectarlas mediante estimación de parámetros del modelo matemático [32].

La técnica consiste en:

- elección de un modelo paramétrico del sistema (en general, modelos lineales con parámetros concentrados)

$$a_n y^n(t) + \dots + a_1 \dot{y}(t) + y = b u(t) + \dots + b_m u^m(t)$$

- determinación de las relaciones entre parámetros del modelo θ_i y parámetros físicos p_i .

$$\theta = f(p)$$

- identificación del vector de parámetros del modelo θ usando la entrada u y la salida y del sistema real.
- determinación del vector de parámetros físicos.

$$p = f^{-1}(\theta)$$

- cálculo del vector de desviaciones Δp de su valor nominal tomado del modelo.
- decisión de falla considerando las relaciones entre variación de los parámetros físicos Δp_i y las fallas.

Este método, particularmente útil en la detección de fallas incipientes, ha sido descrito en detalle por Isermann [32] y Patton *et al.*[46].

3.1.5 El problema de la generación de residuos robustos

Para obtener confiabilidad y buen desempeño en el diagnóstico de fallas, se requiere un modelo que represente con la mayor precisión posible el comportamiento dinámico del sistema. Sin embargo, en los procesos complejos de ingeniería es inevitable la presencia de perturbaciones y errores en el modelado, lo que hace necesario el desarrollo de algoritmos para el diagnóstico robusto de fallas. En este caso, la robustez implica que el sistema debe ser sensible sólo a las fallas, aun en presencia de variación de parámetros, turbulencia, etc. En general, las perturbaciones actúan sobre el proceso en forma incierta, y como consecuencia de ello puede ser difícil diseñar un sistema altamente sensible a las fallas pero insensible frente a perturbaciones no modeladas y con incertidumbre.

Tal como se discutió en la sección anterior, el tema central en el diagnóstico de fallas es la generación de residuos. Estos se ven afectados tanto por las fallas como por las perturbaciones, y discriminar entre ambos efectos es difícil. La clave para el diseño de un sistema robusto de DEF es entonces generar residuos que sean sensibles a las fallas pero insensibles a las incertidumbres y en consecuencia robustos [23], [13].

Analizaremos ahora el modelo en el espacio de estados incluyendo todas las clases posibles de incertidumbre que pueden presentarse, afectando el comportamiento del sistema.

$$\begin{aligned}\dot{x}(t) &= (A + \Delta A)x(t) + (B + \Delta B)u(t) + E_1 d(t) + R_1 f(t) \\ y(t) &= (C + \Delta C)x(t) + (D + \Delta D)u(t) + E_2 d(t) + \\ &\quad + R_2 f(t)\end{aligned}\tag{3.32}$$

donde $d(t) \in \mathbb{R}^q$ es el vector de perturbaciones, E_1 y E_2 son las matrices de distribución de las perturbaciones (se suponen conocidas), R_1 y R_2 son las matrices de distribución de las fallas, y ΔA , ΔB , ΔC , ΔD son matrices que representan errores en los parámetros o variaciones en el modelado.

En cuanto a la descripción del sistema en funciones de transferencia, se obtiene

$$y(s) = (G(s) + \Delta G_u(s))u(s) + G_d(s)d(s) + G_f(s)f(s)\tag{3.33}$$

En esta ecuación, el término $G_d(s)d(s)$ representa el efecto de las perturbaciones y $\Delta G_u(s)$ se utiliza para errores de modelado. Además,

$$G_d(s) = E_2 + C(sI - A)^{-1}E_1$$

Ambos términos $\Delta G_u(s)u(s)$ y $G_d(s)d(s)$ reflejan la incertidumbre en el modelado.

Recordando que $r(s) = H_u(s)u(s) + H_y(s)y(s)$ (ecuación 3.7), y que se debe cumplir la condición (3.8), $G_u(s)H_y(s) + H_u(s) = 0$, al sustituir $y(s)$ en la expresión de $r(s)$, el vector residual en el dominio de la frecuencia queda

$$r(s) = H_y(s)G_f(s)f(s) + H_y(s)\Delta G_u(s)u(s) + H_y(s)G_d(s)d(s)\tag{3.34}$$

Nuevamente se observa que tanto las fallas como las perturbaciones afectan el residuo, y esa es la característica central del problema de robustez en DEF.

3.1.5.1 Robustez frente a perturbaciones

Si el generador de residuos se diseña de modo tal que verifique

$$H_y(s) G_d(s) = 0,$$

la perturbación se encuentra totalmente desacoplada del residuo $r(t)$ y por lo tanto éste es robusto. Este es el principio de desacoplamiento de perturbaciones para la generación de residuos robustos.

Si no se verifica la condición anterior no puede alcanzarse el desacoplamiento total. Podría considerarse un desacoplamiento óptimo o aproximado optimizando un índice de desempeño que contenga una medida de los efectos tanto de las perturbaciones como de las fallas. En el dominio de la frecuencia, una elección posible de dicho índice sería [20]

$$J = \frac{\| H_y(jw) G_f(jw) \|}{\| H_y(jw) G_d(jw) \|} \quad (3.35)$$

Maximizando el índice J sobre un rango especificado de frecuencias puede alcanzarse un diseño con un cierto nivel de desacoplamiento de perturbaciones. Volveremos sobre esto, con más detalle, en el Capítulo 4.

3.1.5.2 Robustez frente a errores de modelado

Para errores de modelado representados por $\Delta G_u(s)$, el problema de la robustez es más complejo de resolver. Se han propuesto dos técnicas principales, la primera conocida como *robustez activa en DEF* y la segunda *robustez pasiva* [24], que utiliza un umbral adaptivo en la etapa de decisión.

La forma *activa* de alcanzar la solución robusta consiste en obtener una estructura aproximada de la incertidumbre, es decir representar errores de modelado como perturbaciones:

$$\Delta G_u(s) \cong G_{d_1}(s) d_1(s)$$

siendo $d_1(s)$ un vector desconocido y $G_{d_1}(s)$ una matriz de transferencia estimada.

En cuanto a la forma *pasiva*, podemos decir que se logra cuando la etapa de decisión es robusta frente a incertidumbres, lo cual implica minimizar la

tasa de falsas alarmas y la pérdida de información sobre fallas que se producen debido al efecto en los residuos de incertidumbres en el modelado y perturbaciones desconocidas.

La robustez pasiva es entonces una alternativa a la robustez activa, y debería ser utilizada cuando la información disponible es muy limitada.

3.2 Generación de residuos robustos utilizando Observadores de Entrada Desconocida (OED)

3.2.1 Teoría y diseño

Como ya se dijo, la generación de residuos robustos es uno de los aspectos más importantes en las técnicas de DEF. En los métodos basados en el desacoplamiento de perturbaciones, se considera que los factores de incertidumbre en el modelado actúan como entradas desconocidas o perturbaciones en un modelo lineal. Si bien el vector de perturbaciones es desconocido, su matriz de distribución se asume conocida. A partir de la información proporcionada por dicha matriz, es posible desacoplar el residuo de las perturbaciones. El principio del OED es, entonces, hacer que el error de estimación de estados no se vea afectado por las entradas desconocidas. Esto ya ha sido estudiado por numerosos investigadores [52], [14], [1].

Comenzaremos a continuación con el diseño de un observador para una clase de sistemas en los cuales la incertidumbre puede ser expresada como un término aditivo de perturbación desconocida en la ecuación dinámica siguiente

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) + Ed(t) \\ y(t) &= Cx(t)\end{aligned}\tag{3.36}$$

donde $x(t) \in \mathbb{R}^n$ es el vector de estados, $y(t) \in \mathbb{R}^m$ es el vector de salida, $u(t) \in \mathbb{R}^r$ es el vector conocido de entrada y $d(t) \in \mathbb{R}^q$ es el vector de perturbaciones. A, B, C, E , son matrices conocidas de dimensiones apropiadas.

3.2.1.1 Algunas observaciones

1. No hay pérdida de generalidad en asumir que la matriz de distribución de las perturbaciones E debe tener rango completo. Si esto no sucede,

puede realizarse la siguiente descomposición [13].

$$E d(t) = E_1 E_2 d(t)$$

donde E_1 es una matriz de rango completo y $E_2 d(t)$ puede considerarse como la nueva entrada desconocida.

2. El término $E d(t)$ puede ser utilizado para describir no sólo una perturbación aditiva, sino también un conjunto de diferentes clases de incertidumbres de modelado, tales como: ruido, términos no lineales en la dinámica del sistema, errores de linealización y reducción del modelo, variaciones de parámetros.

3. El término de perturbación puede aparecer también en la ecuación de salida, tal como:

$$y(t) = C x(t) + E_y d(t)$$

Este puede transformarse en un problema equivalente al anterior mediante una transformación de la señal de salida.

4. Para algunos sistemas, la entrada de control $u(t)$ aparece en la salida, o sea

$$y(t) = C x(t) + D u(t)$$

Siendo la entrada $u(t)$ conocida este problema puede llevarse a uno equivalente al de (3.36) mediante la construcción de una nueva salida.

Definición 3.2.1 *Un observador se define como observador de entrada desconocida para el sistema (3.36) si el vector de error de estimación de estados $e(t)$ tiende a cero asintóticamente, sin ser afectado por la presencia de entradas desconocidas (perturbaciones) en el sistema.*

El problema de diseñar un observador para un sistema lineal con entradas tanto conocidas como desconocidas ha sido estudiado por varios investigadores (en primer término en [61]).

Presentaremos aquí un observador de orden completo, basado en el esquema propuesto en [14] teniendo en cuenta que se requieren grados de

libertad adicionales en el diseño para generar residuos direccionales con el objetivo de aislar fallas.

La estructura del observador queda descrita con las siguientes ecuaciones

$$\begin{aligned}\dot{\hat{z}}(t) &= F z(t) + T B u(t) + K y(t) \\ \hat{x}(t) &= z(t) + H y(t)\end{aligned}\tag{3.37}$$

donde $\hat{x}(t) \in \mathbb{R}^n$ es el vector de estimación de estados, $z(t) \in \mathbb{R}^n$ es el vector de estados de este observador de orden completo y F, T, K, H son matrices a diseñar para lograr el desacoplamiento de perturbaciones y otros objetivos de diseño.

Cuando el observador (3.37) se aplica al sistema (3.36), la dinámica del error de estimación queda definida por la siguiente expresión

$$\begin{aligned}\dot{e}(t) &= (A - HCA - K_1 C) e(t) + [F - (A - HCA - K_1 C)] z(t) \\ &+ [K_2 - (A - HCA - K_1 C) H] y(t) \\ &+ [T - (I - HC)] B u(t) + (HC - I) E d(t)\end{aligned}\tag{3.38}$$

donde $K = K_1 + K_2$ y el error $e(t) = x(t) - \hat{x}(t)$. Si se pueden hacer cumplir las siguientes relaciones

$$(HC - I) E = 0\tag{3.39}$$

$$T = I - HC\tag{3.40}$$

$$F = A - HCA - K_1 C\tag{3.41}$$

$$K_2 = FH\tag{3.42}$$

el error de estimación será, entonces,

$$\dot{e}(t) = F e(t)\tag{3.43}$$

Si todos los autovalores de F son estables el error tenderá a cero asintóticamente, con lo cual el observador (3.37) es un OED para el sistema (3.36), de acuerdo con la definición (3.2.1). En consecuencia, el diseño de este observador consiste en resolver las ecuaciones (3.39)-(3.42), logrando que todos los autovalores de la matriz F del sistema sean estables.

Enunciaremos a continuación dos lemas presentados en [13], y luego analizaremos las condiciones necesarias y suficientes para la existencia de un OED.

Lema 3.2.1 *La ecuación (3.39) tiene solución si y sólo si*

$$\text{rango}(CE) = \text{rango}(E) \quad (3.44)$$

y una solución especial es

$$H^* = E[(CE)^T(CE)]^{-1}(CE)^T \quad (3.45)$$

Lema 3.2.2 *Sea*

$$C_1 = \begin{bmatrix} C \\ CA \end{bmatrix} \quad (3.46)$$

entonces, la detectabilidad del par (C_1, A) es equivalente a la del par (C, A) . Esto es, ambos pares poseen los mismos modos no observables.

Nótese que el par (C, A) es detectable cuando todos sus modos no observables son estables, es decir que la detectabilidad es una condición más débil que la observabilidad.

Teorema 3.2.1 *Las condiciones necesarias y suficientes para que (3.37) sea un OED para el sistema definido por (3.36) son*

- *i) $\text{rango}(CE) = \text{rango}(E)$*
- *ii) (C, A_1) es un par detectable, siendo*

$$A_1 = A - E[(CE)^T(CE)]^{-1}(CE)^T(CA)$$

Demostración:

- Suficiencia: De acuerdo al Lema 3.2.1, la ecuación (3.39) tiene solución si se verifica la condición *i*). Cuando se elige para H la solución especial prevista por dicho lema (H^*), la matriz dinámica del sistema queda

$$F = A - HCA - K_1C = A_1 - K_1C$$

que puede ser estabilizada eligiendo la matriz de ganancia K_1 tal como prevé la condición *ii*). Finalmente, las matrices restantes descritas en (3.37) pueden ser calculadas usando las ecuaciones (3.39)-(3.42). Entonces, el observador (3.37) es un OED para el sistema (3.36).

- Necesidad: Si (3.37) es un OED para el sistema (3.36), la ecuación (3.39) tiene solución. Esto lleva a que se verifique la condición *i*), de acuerdo con el Lema 3.2.1. La solución general de la matriz H para la ecuación (3.39) puede ser calculada como

$$H = E(CE)^+ + H_0[I_m - CE(CE)^+]$$

donde $H_0 \in \mathbb{R}^{n \times m}$ es una matriz arbitraria y $(CE)^+$ es la inversa a izquierda de CE , o sea

$$(CE)^+ = [(CE)^T(CE)]^{-1}(CE)^T$$

Sustituyendo la solución para H en la ecuación (3.41), la matriz dinámica del sistema F es

$$F = A - HCA - K_1C \quad (3.47)$$

$$= A_1 - \overline{K}_1 \overline{C}_1 \quad (3.48)$$

donde $\overline{K}_1 = [K_1 \ H_0]$ y $\overline{C}_1 = \begin{bmatrix} C \\ CA_1 \end{bmatrix}$

Como la matriz F es estable, el par (\overline{C}_1, A_1) es detectable, y también lo es el par (C, A_1) de acuerdo con el Lema (3.2.2).

$\diamond q.e.d.$

3.2.1.2 Observación

Cuando $E = 0$ (no hay perturbaciones en el sistema), tomando $T = I$ y $H = 0$ en (3.37) se obtiene un observador de Luenberger de orden completo. En esta situación, la condición *i*) del Teorema 3.2.1 se verifica claramente, y la condición *ii*) se transforma en la detectabilidad del par (C, A) . Este es un resultado muy conocido en el diseño de un observador de Luenberger.

Del análisis anterior puede concluirse que K_1 es una matriz de parámetros libres en el diseño del OED. Luego de determinarla, otras matrices en el OED pueden ser computadas a partir de las ecuaciones (3.39)-(3.42). La única restricción para la matriz K_1 es que debe estabilizar la matriz de la dinámica del sistema F . La elección de K_1 que estabilice F no es entonces única, debido a la naturaleza multivariable del problema. Es decir, que una vez satisfechas las condiciones impuestas por la existencia de entradas

desconocidas, aún resta alguna libertad para el diseño de K_1 . Esto puede ser utilizado para lograr que el residuo tenga características direccionales o propiedades de mínima varianza (Kalman).

3.3 Esquemas para el aislamiento y detección de fallas basados en OED robustos.

3.3.1 Esquemas robustos para la detección de fallas

Como ya dijimos, el principal objetivo en la detección robusta de fallas es generar un residuo que no se vea afectado por la incertidumbre del sistema. Para detectar una falla en particular, el residuo debe ser sensible a ella. Un sistema con posibles fallas en sensores y actuadores puede describirse así

$$\dot{x}(t) = A x(t) + B u(t) + B f_a(t) \quad (3.49)$$

$$y(t) = C x(t) + f_s(t) \quad (3.50)$$

donde $f_a \in \mathbb{R}^r$ indica la presencia de fallas en los actuadores y $f_s \in \mathbb{R}^m$ indica fallas en los sensores. Para generar un residuo robusto, se necesita un OED como el descrito en la sección anterior. Cuando se dispone de una estimación de los estados, el residuo puede generarse como

$$r(t) = y(t) - C\hat{x}(t) = (I - CH)y(t) - Cz(t) \quad (3.51)$$

Cuando este generador de residuos y el observador (3.37) se aplican al sistema (3.49), el residuo y la dinámica del error de estimación de estados quedan

$$\begin{aligned} \dot{e}(t) &= (A_1 - K_1 C) e(t) + T B f_a(t) - K_1 f_s(t) - H \dot{f}_s(t) \\ r(t) &= C e(t) + f_s(t) \end{aligned} \quad (3.52)$$

De la ecuación (3.52), puede verse que los efectos de las perturbaciones se han desacoplado del residuo. Para detectar fallas en los actuadores, debe hacerse

$$T B \neq 0$$

Más específicamente, la falla en el i -ésimo actuador afectará al residuo si y sólo si

$$T b_i \neq 0$$

donde b_i es la i -ésima columna de la matriz B .

Análogamente, si deben detectarse fallas en los sensores el residuo debe ser sensible frente a $f_s(t)$. Esta condición se verifica en general, ya que el vector de fallas en los sensores tiene un efecto directo sobre el residuo.

El residuo robusto puede ser utilizado para detectar fallas mediante

$$\|r(t)\| = \begin{cases} < \Lambda, & \text{si no hay fallas} \\ \geq \Lambda, & \text{si hay fallas} \end{cases}$$

donde Λ es el umbral.

3.3.2 Esquemas robustos para el aislamiento de fallas

El problema del aislamiento de la falla consiste en ubicarla, es decir, determinar en qué sensor o actuador ha sucedido. Como se dijo en la Sección 3.1.3, uno de los posibles métodos para lograr el aislamiento de la falla consiste en diseñar un conjunto de residuos *estructurado*. En este caso “estructurado” significa que cada residuo se diseña para ser sensible frente a cierto grupo de fallas e insensible frente a otras. Las propiedades de sensibilidad e insensibilidad hacen posible el aislamiento. La situación ideal es hacer cada residuo sólo sensible a una falla en particular e insensible a todas las otras. Sin embargo, esta situación ideal es en general difícil de alcanzar, y aunque pudiera lograrse, no quedarían grados de libertad en el diseño para conseguir la robustez. Para explotar al máximo la libertad de diseño en función de la robustez, un esquema comúnmente aceptado [46] se basa en hacer cada residuo sensible a fallas en todos los sensores o actuadores menos en uno.

3.3.2.1 Fallas en sensores

Para diseñar esquemas robustos para el aislamiento de fallas en sensores, todos los actuadores se suponen libres de falla y las ecuaciones del sistema pueden expresarse como

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) + Ed(t) \\ y^j(t) &= C^j x(t) + f_s^j(t) \quad \text{para } j = 1, 2, \dots, m \\ y_j(t) &= c_j x(t) + f_{sj}(t) \end{aligned} \tag{3.53}$$

donde $c_j \in \mathbb{R}^{1 \times n}$ es la j -ésima columna de la matriz C , $C^j \in \mathbb{R}^{(m-1) \times n}$ se obtiene de la matriz C al eliminar la j -ésima columna c_j , y_j es la j -ésima componente de y , e $y^j \in \mathbb{R}^{m-1}$ se obtiene del vector y eliminando la j -ésima componente y_j .

Partiendo de esta descripción, el generador de residuos puede construirse como

$$\begin{aligned}\dot{z}(t) &= F^j z^j(t) + T^j B u(t) + K_j y_j(t) \\ r_j(t) &= (I - C_j H_j) y^j(t) - C^j z^j(t)\end{aligned}\quad (3.54)$$

para $j = 1, 2, \dots, m$, donde las matrices de parámetros deben satisfacer las siguientes ecuaciones

$$\begin{aligned}H^j C^j E &= E \\ T^j &= I - H^j C^j \\ F^j &= T^j A - K_1^j C^j \\ K_2^j &= F^j H^j \\ K^j &= K_1^j + K_2^j\end{aligned}\quad (3.55)$$

para $j = 1, 2, \dots, m$, y donde F^j debe ser estabilizada. Se observa que cada generador de residuos recibe todas las entradas y todas menos una de las salidas. Cuando no hay fallas en los actuadores y sucede una falla en el j -ésimo sensor, el residuo satisface la siguiente ley lógica

$$\begin{aligned}\|r^j(t)\| &< \Lambda^j \\ \text{y } \|r^k(t)\| &\geq \Lambda^k\end{aligned}$$

para $k = 1, 2, \dots, j-1, j+1, \dots, m$ siendo Λ^j $j = 1, 2, \dots, m$ los umbrales de aislamiento.

3.3.2.2 Fallas en actuadores

En este caso se supone que no hay fallas en los sensores, y las ecuaciones del sistema quedan

$$\begin{aligned}\dot{x}(t) &= Ax(t) + B^i u^i(t) + E^i d^i(t) \\ y(t) &= Cx(t)\end{aligned}\quad (3.56)$$

para $i = 1, 2, \dots, r$, y donde b_i es la i -ésima columna de la matriz B , $B^i \in \mathbb{R}^{n \times (r-1)}$ se obtiene de la matriz B anulando la i -ésima columna b_i , u_i es la i -ésima componente de u , $u^i \in \mathbb{R}^{(r-1)}$ se obtiene del vector u eliminando la i -ésima componente u_i , y

$$E^i = \begin{bmatrix} E & b_i \end{bmatrix} \quad d^i(t) = \begin{bmatrix} d(t) \\ u_i(t) + f_{ai}(t) \end{bmatrix}$$

para $i = 1, 2, \dots, r$. Entonces, los OED generadores de residuos pueden ser contruidos así

$$\begin{aligned} \dot{z}^i(t) &= F^i z^i(t) + T^i B^i u^i(t) + K^i y(t) \\ r_i(t) &= (I - C H^i) y(t) - C z^i(t) \end{aligned} \quad (3.57)$$

para $i = 1, 2, \dots, r$. Las matrices de parámetros deben satisfacer las siguientes ecuaciones

$$\begin{aligned} H^i C^i E &= E^i \\ T^i &= I - H^i C \\ F^i &= T^i A - K_1^i C \\ K_2^i &= F^i H^i \\ K^i &= K_1^i + K_2^i \end{aligned} \quad (3.58)$$

Se observa que cada generador de residuos tiene todas las salidas y todas menos una de las entradas. Cuando no hay fallas en los sensores y ocurre una falla en el i -ésimo actuador, el residuo cumple la siguiente ley lógica

$$\begin{aligned} \| r^i(t) \| &< \Lambda^i \\ \text{y } \| r^k(t) \| &\geq \Lambda^k \end{aligned}$$

para $k = 1, 2, \dots, i-1, i+1, \dots, m$; siendo Λ^i $i = 1, 2, \dots, r$ los umbrales de aislamiento.

3.3.2.3 Observaciones

Los esquemas presentados pueden aislar una sola falla por vez, ya sea en un sensor o en un actuador. Esto se fundamenta en que la probabilidad de ocurrencia de dos o más fallas simultáneas es muy baja en un caso real. Si se necesita aislar dos o más fallas al mismo tiempo, el esquema de aislamiento debería modificarse mediante un reagrupamiento de fallas. El modo de agrupamiento depende del sistema y de las fallas a ser aisladas, lo cual reafirma

el hecho de que en DEF no puede esperarse un esquema general aplicable a cualquier sistema, sino que por el contrario cada caso debe analizarse en forma particular.

3.4 Detección de fallas en sistemas no lineales

Cuando el sistema en estudio es fuertemente no lineal, el uso de aproximaciones lineales se ve limitado. El análisis de la DEF en sistemas no lineales juega un papel muy importante en las aplicaciones prácticas, ya que muchos procesos industriales son de naturaleza no lineal y, en consecuencia, su modelo matemático debe reflejarlo.

Dado que no es de esperar una simple extensión de los resultados existentes en DEF en el campo lineal al no lineal, fue necesario desarrollar métodos que ataquen directamente el problema de la detección de fallas para los modelos no lineales. Con la aplicación de la teoría de los observadores no lineales (ONL) se han obtenido resultados principalmente en detección y, con algunas restricciones, también en el aislamiento de fallas. Algunos problemas, que consideran modelos más generales así como el diseño de los ONL correspondientes, se encuentran aún abiertos debido a las dificultades para estimar los estados o el vector de mediciones en un SNL, incluso en el caso en que las no linealidades son conocidas y no hay perturbaciones [1].

Presentaremos en lo que sigue las principales técnicas para generación de residuos en SNL, y analizaremos luego el problema de la robustez en estos sistemas, considerando el rechazo o atenuación de perturbaciones en ellos.

Un modelo no lineal puede representarse con el siguiente sistema de ecuaciones

$$\dot{x}(t) = g(x(t), u(t), f(t), d(t)) \quad (3.59)$$

$$y(t) = h(x(t), u(t), f(t), d(t)) \quad (3.60)$$

donde $x(t)$ es el vector de estados, $u(t)$ es el vector de entrada, $y(t)$ es el vector de salida, $f(t)$ es el vector de fallas, $d(t)$ es el vector de perturbaciones y $g(\cdot, \cdot, \cdot, \cdot)$ y $h(\cdot, \cdot, \cdot, \cdot)$ indican transformaciones no lineales (funciones). El problema de la DEF consiste entonces en generar un vector residual $r(t)$ utilizando la siguiente estructura

$$\dot{\xi}(t) = g_r(\xi(t), u(t), y(t)) \quad (3.61)$$

$$r(t) = h_r(\xi(t), u(t), y(t)) \quad (3.62)$$

El principal desafío se encuentra en el diseño de las transformaciones no lineales g_r y h_r .

3.4.1 Observadores no lineales para generación de residuos

3.4.1.1 Observador identidad no lineal

Análogamente al sistema lineal, esta técnica se basa en el diseño de un observador identidad para el sistema controlado. Consideremos un modelo no lineal como el que sigue

$$\dot{x}(t) = g(x(t), u(t)) + R_1 f(t) \quad (3.63)$$

$$y(t) = h(x(t), u(t)) + R_2 f(t) \quad (3.64)$$

Recordemos que R_1 y R_2 son las matrices de distribución de las fallas.

Un observador identidad no lineal para este sistema puede diseñarse así

$$\dot{\hat{x}}(t) = g(\hat{x}(t), u(t)) + K(\hat{x}(t), u(t)) [y(t) - \hat{y}(t)] \quad (3.65)$$

$$\hat{y}(t) = h(\hat{x}(t), u(t)) \quad (3.66)$$

$$r(t) = y(t) - \hat{y}(t) \quad (3.67)$$

El residuo $r(t)$ y la dinámica del error de observación quedan

$$\begin{aligned} \dot{e}(t) &= F(t)e(t) + O(e^2(t), t) + R_1 f(t) - \\ &\quad - K(\hat{x}(t), u(t)) R_2 f(t) \end{aligned} \quad (3.68)$$

$$r(t) = H(t)e(t) + O(e^2(t), t) + R_2 f(t) \quad (3.69)$$

donde $O(e^2(t), t)$ representa los términos de segundo y mayor orden con respecto a $e(t)$. También se tiene

$$F(t) = \frac{\partial g(\hat{x}(t), u(t))}{\partial \hat{x}(t)} - K(\hat{x}(t), u(t)) H(t) \quad (3.70)$$

$$H(t) = \frac{\partial h(\hat{x}(t), u(t))}{\partial \hat{x}(t)} \quad (3.71)$$

En lo anterior puede verse que si el error converge asintóticamente a cero el residuo sólo se ve afectado por la falla. El problema reside entonces en diseñar una matriz $K(\hat{x}(t), u(t))$ de modo que el origen sea un equilibrio asintóticamente estable de la ecuación (3.68). En muchas situaciones prácticas puede elegirse una matriz constante K [25].

Debe remarcar que no se ha hallado un algoritmo general que resuelva este problema de estabilización, ya que pueden aparecer dificultades computacionales y numéricas debido a la no linealidad del modelo.

3.4.1.2 Observador de Thau

Este observador fue desarrollado por Thau para una clase especial de sistemas no lineales, tales como

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) + R_1 f(t) + g(x(t), u(t)) \\ y(t) &= Cx(t) + R_2 f(t)\end{aligned}\tag{3.72}$$

El modelo satisface las siguientes condiciones

- el par (C, A) es observable.
- la función no lineal $g(x(t), u(t))$ es continuamente diferenciable y localmente Lipschitz con constante ρ , es decir

$$\|g(x_1, u) - g(x_2, u)\| \leq \rho \|x_1 - x_2\|$$

Cuando se satisfacen estas condiciones, un observador estable para el sistema (3.74) se construye así

$$\begin{aligned}\dot{\hat{x}}(t) &= A\hat{x}(t) + Bu(t) + g(\hat{x}(t), u(t)) + K(y(t) - \hat{y}(t)) \\ \hat{y}(t) &= C\hat{x}(t)\end{aligned}\tag{3.73}$$

donde K es la matriz de ganancia del observador

$$K = P_{\Theta}^{-1} C^T$$

La matriz P_{Θ} es la solución de la ecuación de Lyapunov [50]

$$A^T P_{\Theta} + P_{\Theta} A - C^T C + \Theta P_{\Theta} = 0\tag{3.74}$$

que también puede escribirse

$$(A + \frac{\theta}{2} I_n)^T P_{\Theta} + P_{\Theta} (A + \frac{\theta}{2} I_n) - C^T C = 0$$

en la cual Θ es un parámetro positivo elegido de modo que la solución de dicha ecuación sea positiva definida.

Cabe destacar que la versión del observador de Thau que aquí mostramos se debe a Schreier [50] y conduce a la ecuación (3.74). El observador planteado originalmente por Thau lleva, por su parte, a otro tipo de ecuación de Lyapunov en la cual las matrices A no son constantes.

3.4.1.3 Observador no lineal de entrada desconocida

En principio, la idea presentada en la Sección 3.2 para los OED lineales fue extendida a una clase especial de sistemas no lineales por Wünnenberg [65]. Estos sistemas pueden ser modelados como sigue

$$\begin{aligned}\dot{x}(t) &= Ax(t) + B(y(t), u(t)) + E_1 d(t) + R_1 f(t) \\ y(t) &= Cx(t) + E_2 d(t) + R_2 f(t)\end{aligned}\quad (3.75)$$

donde $f(t)$ es el vector de fallas y $d(t)$ el vector de perturbaciones. Notemos que el término no lineal $B(y(t), u(t))$ sólo depende de la salida y y la entrada u , que se encuentran disponibles en forma directa. En consecuencia, la no linealidad puede ser compensada completamente utilizando un observador de la forma

$$\begin{aligned}\dot{\xi}(t) &= F\xi(t) + J(y(t), u(t)) + Gy(t) \\ r(t) &= L_1\xi(t) + L_2y(t)\end{aligned}\quad (3.76)$$

Para proveer robustez frente a perturbaciones y sensibilidad a las fallas, las matrices del observador deben cumplir con las siguientes condiciones

$$\begin{aligned}TA - FT &= GC && F \text{ estable} \\ J(y, u) &= TB(y, u) \\ GE_2 - TE_1 &= 0 \\ L_2E_2 &= 0 \\ L_1T + L_2C &= 0 \\ \text{rango}(GR_2 - TR_1) &= \text{rango}(R_1) \\ \text{rango}(L_2R_2) &= \text{rango}(R_2)\end{aligned}\quad (3.77)$$

Si se cumplen esas condiciones, la dinámica del error de estimación, $e(t) = \xi(t) - Tx(t)$, y el residuo $r(t)$ son

$$\begin{aligned}\dot{e}(t) &= Fe(t) + GR_2f(t) - TR_1f(t) \\ r(t) &= L_1e(t) + L_2R_2f(t)\end{aligned}\quad (3.78)$$

El inconveniente que presenta esta extensión de la teoría de los OED lineales a una clase de sistemas no lineales es, justamente, la limitación de que sólo pueda aplicarse a modelos como los descritos en (3.75) y que se deba cumplir con las ecuaciones (3.77).

Una técnica alternativa, que requiere condiciones de existencia más débiles y extiende la clase de sistemas a un modelo más general, fue propuesta en [52]

$$\begin{aligned}\dot{x}(t) &= A(x) + B(x)u + E(x)d + R(x)f \\ y(t) &= Cx(t)\end{aligned}\tag{3.79}$$

El objetivo del diseño es hallar una transformación no lineal $\xi = T(x)$ que desacople las perturbaciones en el modelo. Esto puede lograrse si y sólo si

$$\frac{\partial T(x)}{\partial x} E(x) = 0\tag{3.80}$$

Esta relación constituye un sistema de ecuaciones diferenciales en derivadas parciales que debe ser resuelto simultáneamente a $\xi = T(x)$. Suponiendo que existan dichas soluciones, el modelo puede reescribirse como sigue

$$\dot{\xi} = \frac{\partial T(x)}{\partial x} (A(x) + B(x)u + R(x)f)\tag{3.81}$$

donde la transformación de la salida $y^* = C^*(y)$ denota un subconjunto de las mediciones disponibles $y = C(x)$ sujeto a la condición

$$\dim(y^*) < \dim(y)$$

Supongamos además que existe una relación $Q(T(x), C(x)) = 0$. Esta condición es limitante puesto que en una cierta subvariedad, el conjunto de salidas debe tener una relación funcional especial con $\xi(x)$. Es decir:

$$y = Cx = CT^{-1}(\xi) \rightarrow y - CT^{-1}(\xi) = 0 = Q(\xi, y)$$

Entonces un ONL puede estimar la porción “no perturbada” ξ del estado x . El observador resultante tiene la forma

$$\dot{\hat{\xi}} = \frac{\partial T(\hat{x})}{\partial \hat{x}} \left(A(\hat{x}) + B(\hat{x})u + K(\hat{\xi}, y, u)Q(\hat{\xi}, y) \right)\tag{3.82}$$

Aquí, la elección de la matriz de realimentación $K(\hat{\xi}, y, u)$ debe realizarse en función de estabilizar la ecuación diferencial que rige el error de estimación $e = \hat{\xi} - \xi$. Las relaciones $Q(\hat{\xi}, y)$ pueden utilizarse como residuos.

$$r = Q(\hat{\xi}, y) = Q(\xi + e, y)\tag{3.83}$$

La dinámica del error de estimación es

$$\dot{e}(t) = \rho(e, t) - \frac{\partial T(x)}{\partial x} R(x) f \quad (3.84)$$

donde se desea que el punto de equilibrio $e = 0$ sea al menos local y asintóticamente estable. Entonces, el residuo convergerá a cero si no hay fallas. Por otra parte, todas las fallas en el vector f se reflejarán en el error e si

$$\text{rango} \left[\frac{\partial T(x)}{\partial x} R(x) \right] = \text{rango} [R(x)]$$

3.4.1.4 Observador adaptivo no lineal

Consideremos el sistema no lineal descrito en [20]

$$\begin{aligned} \dot{x} &= a(x) + q_0(x, u) + Q(x, u)\theta + G(x, u)f + g(t) \\ y &= c(x) \quad x_0 = x(0) \end{aligned} \quad (3.85)$$

donde la salida y se considera escalar por simplicidad, $a : \mathbb{R}^n \rightarrow \mathbb{R}^n$, $Q : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{n \times l_d}$, $c : \mathbb{R}^n \rightarrow \mathbb{R}$, $g : \mathbb{R}^+ \rightarrow \mathbb{R}^n$, $G : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{n \times l}$, se suponen conocidas y suficientemente suaves, f representa cambios abruptos y $\theta \in \mathbb{R}^{l_d}$ es un vector desconocido que representa, por ejemplo, parámetros desconocidos variantes en el tiempo, fallas lentamente variables o parte de las no linealidades del sistema. Se asume que $0 < \|\dot{\theta}\| \leq M \ll \infty$.

El generador de residuos adaptivo se diseña en dos pasos [1]

1. Se halla, si es posible, una transformación $\xi = T(x) \in \mathbb{R}^k$, $k < n$, definida en un entorno del estado inicial x_0 de modo que

$$\begin{aligned} \dot{\xi} &= F\xi + \psi_0(y, u) + \Psi(y, u)\theta + \Xi(x, u)f \\ y &= [0 \dots 0 \ 1] \xi \end{aligned} \quad (3.86)$$

y el rango de $\Xi(y, u) = l$.

2. Se diseña el residuo adaptivo como

$$\begin{aligned} \dot{z} &= Fz + \psi_0(y, u) + \Psi(y, u)\hat{\theta} + L_1 y + \begin{bmatrix} V(t) \\ 0 \end{bmatrix} \dot{\hat{\theta}} \\ r(t) &= y(t) - z_k(t) \\ \hat{\theta}(t) &= \Gamma \phi^T(t) r(t) \\ \dot{V}(t) &= RV(t) + \Psi_k(y, u) \quad V(0) = 0 \\ \phi(t) &= k^T V(t) + \Psi_k(y, u) \end{aligned} \quad (3.87)$$

donde Γ es una matriz definida positiva, R es una matriz estable, $k^T = [0 \dots 0 \ 1]$, y los elementos l_j de L_1 se eligen de modo que $F - L_1 C$ sea Hurwitz.

Para garantizar que $|y - z_k| < K < \infty$, deben cumplirse además las siguientes condiciones:

- $\phi(t)$ es acotada, $\dot{\phi}(t)$ es acotada excepto en un número contable de puntos.
- $\exists \alpha, \beta$, tales que $0 < \alpha I \leq \int_t^{t+\beta} \phi(\tau) \phi^T(\tau) d\tau$
- $\exists M_1$ tal que $|V(t)\dot{\theta}| \leq M_1 < \infty$

Se asume en general que $\|\dot{\theta}\|$ es pequeño. Entonces, es razonable suponer que los errores de estimación sean cercanos a cero.

Otra técnica relacionada con ésta se basa en el observador propuesto en [28]. Los sistemas considerados deben ser observables para toda entrada, en el sentido propuesto en [27], y transformables en la forma

$$\begin{aligned} \dot{x} &= F'(x) + G'(x)u + T'(y, u)d + p(x)f \\ y &= Cx \end{aligned} \quad (3.88)$$

donde:

$$\begin{aligned} F'(x) &= [x_2, \dots, x_n, \phi(x)]^T \\ G'(x) &= [g_1(x_1), g_2(x_1, x_2), \dots, g_n(x_1, \dots, x_n)]^T \\ C &= [1 \ 0 \dots 0] \\ T'(y, u) &= S_\infty^{-1} \begin{bmatrix} \tau_1 & \dots & \tau_s \\ 0 & \dots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \dots & 0 \end{bmatrix} \end{aligned}$$

Aquí, $d \in \mathbb{R}^s$ es un vector de perturbaciones no medibles que representa, por ejemplo, parámetros lentamente variables en el tiempo o parte de las no linealidades del sistema. S_∞ es la solución de la ecuación de Lyapunov:

$$-\gamma S_\infty - A^T S_\infty - S_\infty A + C^T C = 0$$

siendo A la matriz de Brunovsky (todas las entradas ceros y sólo unos en la primer superdiagonal).

En el caso en que $d = f = 0$, la observabilidad garantiza la existencia de la transformación [28].

Bajo la hipótesis de que las funciones g_i son Lipschitz, un observador adaptivo para el sistema anterior viene dado por

$$\begin{aligned}\dot{z} &= F'(z) + G'(z)u + T'(y, u)\hat{\theta}_d + S_\infty^{-1}C^T(z_1 - y) \\ \dot{r} &= z_1 - y \\ \dot{\hat{\theta}}_d &= -[\tau_1 \dots \tau_s]^T [z_1 - y]\end{aligned}\tag{3.89}$$

Para entradas u uniformemente acotadas por algún $u_0 \geq 0$, el sistema (3.89) es un observador de (3.88), es decir, para θ suficientemente grande

$$\|z(t) - x(t)\| \leq K(\theta) \exp(-\theta t/3) \|z_0 - x_0\|$$

Puede verse que el observador propuesto es sencillo, ya que copia la dinámica del sistema, agregando un término correctivo que sólo depende de la dimensión y la velocidad deseada de convergencia θ . Como converge para valores de θ suficientemente grandes, suele denominarse *Observador de alta ganancia*, ya que a mayor θ la velocidad de convergencia también es mayor. Notemos que la consecuencia no deseable de una ganancia elevada es la amplificación del ruido.

3.4.1.5 Filtro detector de fallas no lineal

Presentaremos una extensión del filtro detector de fallas para sistemas lineales a una clase de sistemas no lineales [1]. Sea entonces un sistema no lineal representable por

$$\begin{aligned}\dot{x} &= Ax + f(x) + Bu + \sum_{i=1}^s F_i \phi_i(x, u) f_i \\ y &= Cx\end{aligned}\tag{3.90}$$

donde $F_i \in \mathbb{R}^n$, s es la cantidad de cambios en los parámetros considerada (fallas $f_i = \theta_{fi} - \theta_{f0i}$) y ϕ_i es una función dependiente de los parámetros considerados.

Las hipótesis requeridas son las que siguen.

- el término no lineal es Lipschitz, o sea

$$\| f(x_1) - f(x_2) \| \leq \gamma \| x_1 - x_2 \|$$

- (A, C) es observable
- $CF = [CF_1 \ CF_2 \ ... \ CF_s]$ tiene rango s , es decir, los modos de falla son independientes en el espacio de salida.

Tal como en el caso lineal, los filtros de detección son observadores de la salida que producen residuos con propiedades direccionales. Consideremos un observador de tipo Thau para el sistema (3.90)

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + Bu + f(\hat{x}) + H(y - C\hat{x}) \\ r &= y - C\hat{x}(t) \end{aligned} \tag{3.91}$$

donde la matriz H tiene la siguiente expresión

$$H = Q_h[CF]^\dagger + H_F[I - CF(CF)^\dagger]C$$

siendo Q_h una matriz que depende de los autovalores y autovectores de $A_H = A - HC$, H_F está relacionada con la solución de una ecuación de Riccati modificada, y el supraíndice \dagger indica la pseudoinversa.

En síntesis, el objetivo de diseño en un filtro detector de fallas consiste en hacer que el residuo sea unidireccional, eligiendo adecuadamente la matriz de ganancia H . Puede observarse que en este análisis no han sido considerados factores de incertidumbre, tales como perturbaciones o errores de modelado. De hecho, esta es la principal desventaja de estos filtros dado que, como ya se dijo, es inevitable que aparezcan elementos inciertos en un sistema real.

3.5 Otros enfoques para el análisis de la detección de fallas

Los diferentes métodos desarrollados muestran propiedades diversas con respecto al diagnóstico de fallas distintas en un proceso, por lo cual una integración de varias técnicas puede ser útil para lograr un diagnóstico confiable [32].

Por otra parte, se han desarrollado últimamente técnicas que utilizan la lógica difusa o las redes neuronales para la detección y aislamiento de fallas.

3.5.1 Lógica difusa en el diagnóstico de fallas

La segunda etapa de la DEF, es decir la toma de decisión, es un proceso lógico que transforma conocimiento cuantitativo (señales residuales) en sentencias cualitativas (normal, con fallas, etc.). Esto hace suponer que podría tratarse con la ayuda de lógica difusa. Consideremos que el residuo debido a las fallas está también contaminado con ruido y el efecto de incertidumbre por desacoplamiento incompleto de perturbaciones, con lo cual será distinto de cero aun en ausencia de fallas. Típicamente estos efectos serán variantes en el tiempo, por lo que el residuo fluctuará dependiendo de funciones desconocidas del tiempo y de las perturbaciones, ruido y entradas del proceso. Esta es una situación común, y en consecuencia la lógica difusa parece ser una herramienta natural para la etapa de decisión, en una situación incierta y compleja basada en información incompleta.

Contrariamente a la lógica clásica, que sólo permite una clasificación definida con valores fijos, la lógica difusa ofrece una forma de descripción de las tolerancias, es decir, valores difusos, reglas heurísticas y su combinación, aunque con mucha menor precisión que con un modelo probabilístico. Si bien el método posibilita la toma de decisión y provee una DEF aplicable a sistemas industriales, la dificultad se encuentra en el entrenamiento del algoritmo en el mecanismo de inferencia [13].

3.5.2 Redes neuronales en el diagnóstico de fallas en SNL

Una red neuronal puede utilizarse para modelar sistemas dinámicos no lineales con múltiples entradas y salidas. Luego del entrenamiento, la red puede proporcionar una estimación ajustada de la salida del sistema. Tomando el concepto de generación de residuos analizado en secciones precedentes, la diferencia pesada entre salidas reales y estimadas se usa como residuo para detectar fallas. Cuando la magnitud de esta diferencia excede un umbral preestablecido, puede concluirse que el sistema ha fallado.

Para localizar fallas en el sistema en forma confiable, se dispone una red neuronal secundaria para examinar características del residuo. Una ca-

racterística particular corresponde a una ubicación específica de la falla. Basada en la extracción de características y principios de clasificación, la segunda red puede aislar fallas de manera confiable. En Chen y Patton [13] se presenta una técnica que utiliza redes neuronales en todas las etapas de la detección y aislamiento de fallas.

Si bien se presentan allí resultados de simulación que demuestran que un esquema basado en redes neuronales pueden diagnosticar fallas de forma confiable en sistemas no lineales, existe un límite para la detección de fallas pequeñas porque no es posible entrenar a la red neuronal cuando la magnitud de la señal de falla es comparable con el ruido y errores de modelado. Por otra parte, aún no se ha analizado la robustez del esquema frente a ruido y perturbaciones. Un buen esquema de DEF debería desempeñarse correctamente incluso cuando las condiciones de operación hubieran cambiado luego del entrenamiento, en cuyo caso una posibilidad es entrenar la red periódicamente.

Cabe destacar, finalmente, que un modelo probabilístico también puede “entrenarse” (ajustarse), proveyendo reglas de decisión racionales con hipótesis explícitas, lo cual constituye una importante ventaja frente a técnicas como redes neuronales o lógica difusa, cuya utilización sólo se justifica en modelos con una notable incertidumbre.

Capítulo 4

Desacoplamiento o atenuación de perturbaciones en sistemas de detección de fallas

Tal como se vio en el capítulo anterior, para que el sistema de DEF sea confiable los efectos de las perturbaciones o entradas desconocidas al sistema deben poder desacoplarse de las fallas. Dicho de otra manera, el residuo no debería reflejar la influencia de las perturbaciones. Esto no siempre es fácil de lograr en los sistemas reales, y menos aún si el modelo es no lineal.

Analizaremos en este capítulo el problema del desacoplamiento total para el caso de sistemas lineales y la alternativa del desacoplamiento aproximado, que consiste en la optimización de índices de desempeño en el caso en que no se verifiquen ciertas condiciones necesarias para el desacoplamiento total.

En cuanto a los sistemas no lineales, estudiaremos también la posibilidad de rechazo de perturbaciones y, cuando esto no sea factible, plantearemos el problema de atenuación de las mismas. Observemos que mientras en [33] se desarrolla este problema para controladores, el planteo que aquí proponemos es nuevo, y equivale a lo que en sistemas lineales sería el dual del anterior: la atenuación de perturbaciones para observadores no lineales. Además, extendaremos el análisis al caso multivariable (perturbación no escalar), que no ha sido tratado en [33].

4.1 Desacoplamiento de perturbaciones en sistemas lineales

Una descripción típica de las incertidumbres del sistema se basa, tal como se vio en el capítulo anterior, en el concepto de *entradas desconocidas* actuando sobre un modelo lineal nominal:

$$\begin{aligned}\dot{x}(t) &= Ax(t) + Bu(t) + Ed(t) + R_1 f(t) \\ y(t) &= Cx(t) + Du(t) + R_2 f(t)\end{aligned}\tag{4.1}$$

donde el término de perturbación $Ed(t)$ representa incertidumbres que actúan sobre el sistema, $d(t) \in \mathbb{R}^q$ es el vector de perturbaciones o entradas desconocidas, y la matriz de distribución $E \in \mathbb{R}^{n \times q}$ se asume conocida. En la DEF robusta, esta descripción de la incertidumbre del sistema se conoce como *incertidumbre estructurada*.

En la ecuación (4.1) puede verse que los términos $Ed(t)$ y $R_1 f(t)$ actúan en el sistema de la misma forma, por lo cual no es posible discriminar sus efectos si no se conoce la estructura de E . En general se asume en la práctica que ésta es conocida, y entonces el residuo puede construirse de modo que posea la propiedad de desacoplamiento robusto de perturbaciones. Luego, este residuo robusto puede usarse para detectar fallas en forma confiable.

En la Sección 3.2 hemos estudiado la teoría y diseño de los *Observadores de entrada desconocida* (OED) cuyo objetivo es, precisamente, la generación de residuos robustos para modelos lineales. Vimos allí, además, algunas condiciones necesarias y suficientes para su existencia teniendo en cuenta que el error de estimación de estados no debe verse afectado por las perturbaciones. En síntesis, dicho análisis presenta un panorama del problema de desacoplamiento de perturbaciones en sistemas lineales en el dominio del tiempo. Desarrollaremos entonces otro aspecto, que consiste en el estudio del desacoplamiento de perturbaciones en el dominio de la frecuencia.

4.1.1 Desacoplamiento de perturbaciones en el dominio de la frecuencia

Como vimos en la Sección 3.1.4, los observadores lineales para la DEF pueden diseñarse como observadores de la salida del sistema. Una forma directa de construir un observador de la salida consiste en utilizar la relación

entrada-salida descrita por una función de transferencia en el dominio de la frecuencia, tal como se propone en [23].

Consideremos un proceso lineal cuya salida puede describirse

$$y(s) = G_u(s)u(s) + \Delta y(s) + G_f f(s) \quad (4.2)$$

donde G_u , G_f son matrices de transferencia conocidas del vector de entrada $u \in \mathbb{R}^p$ y del vector de fallas $f \in \mathbb{R}^q$ al vector de salida $y \in \mathbb{R}^m$. $\Delta y(s)$ es un vector que representa perturbaciones desconocidas e incertidumbre en el modelado.

El generador de residuos puede realizarse por la técnica de factorización, que es aplicable tanto a procesos estables como inestables [20].

$$r(s) = Q(s) [\hat{M}_u(s)y(s) - \hat{N}_u(s)u(s)] \quad (4.3)$$

donde $\hat{M}_u(s)$ y $\hat{N}_u(s)$ son los factores coprimos a izquierda de $G_u(s)$ y satisfacen $\hat{M}_u^{-1}(s)\hat{N}_u(s) = G_u(s)$ (véase la Figura 4.1). Por su parte, $Q \in RH_\infty$ ¹ es una matriz de parámetros libres a seleccionar. Su significado físico es el de un posfiltro que provee grados adicionales de libertad en el diseño con el objeto de, por ejemplo, aislar fallas o generar residuos robustos.

Cabe destacar que un trabajo reciente [30] extiende esta técnica al diseño de sistemas no lineales de control, definiendo para ellos la *factorización dinámica a derecha* con el objeto de analizar la observabilidad y estudiar el problema de diseño de un observador para SNL.

Para analizar la robustez del residuo, debemos estudiar la influencia que sobre él ejercen las fallas e incertidumbres modeladas con Δy . Sustituyamos con este fin la ecuación (4.2) en (4.3), para obtener

$$r(s) = Q(s)\hat{M}_u(s)[G_f(s)f(s) + \Delta y(s)] \quad (4.4)$$

La incertidumbre del modelo Δy puede ser estructurada o no estructurada, pero a los efectos de la detección de fallas sólo se necesita considerar la siguiente forma general

$$\Delta y(s) = G_d(s)d(s),$$

¹ RH_∞ es un subespacio de $\mathcal{L}_\infty(j\mathbb{R}, \mathbb{C}^{n \times m})$ de funciones racionales de transferencia o matrices de transferencia, analíticas y acotadas en $\operatorname{Re}(s) > 0$.

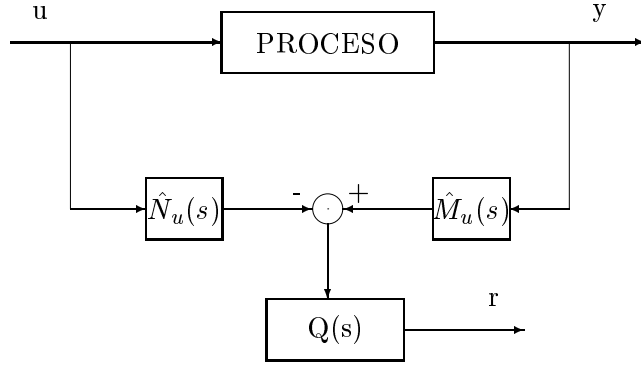


Figura 4.1: Generación de residuos mediante factorización

con un vector d desconocido pero acotado, de modo que $\|d\|_2 < \delta_d$.

Si el rango de $G_d(s) = m$ y no se dispone de información en el dominio de la frecuencia, por ejemplo $G_d(s) = I$, la expresión $G_d(s)d(s)$ representa incertidumbre no estructurada y de otro modo, incertidumbre estructurada. Entonces, de la ecuación (4.4) se sigue que

$$r(s) = Q(s)\hat{M}_u(s)[G_f(s)f(s) + G_d(s)d(s)] \quad (4.5)$$

Esta última es la forma usual en que se expresa el generador de residuos para determinar la matriz de parametrización $Q(s)$ que cumpla con las especificaciones deseadas.

4.1.1.1 Desacoplamiento total

Para el aislamiento perfecto de las fallas y la invariancia total frente a entradas desconocidas $d(s)$ se requiere desacoplamiento total no sólo entre fallas, sino también entre fallas y perturbaciones.

Esto último sólo puede lograrse si la incertidumbre es estructurada, en cuyo caso y de acuerdo con la ecuación (4.5) la matriz $Q(s)$ debe verificar las siguientes condiciones

$$\begin{aligned} Q(s)\hat{M}_u(s)G_f(s) &= \text{diag}(t_1(s), \dots, t_q(s)) \in RH_\infty \\ Q(s)\hat{M}_u(s)G_d(s) &= 0 \end{aligned} \quad (4.6)$$

con t_i arbitrarios.

Con respecto al filtro $Q(s)$, desde el punto de vista físico éste permite utilizar las diferentes frecuencias características de los efectos de las fallas y entradas desconocidas como un criterio adicional para su discriminación. Además, en forma práctica el residuo generalizado de forma (4.3) puede verse como una extensión de las técnicas de los espacios de paridad y de los observadores robustos en el dominio del tiempo. Si bien se ha demostrado que existen casos prácticos en donde puede lograrse el desacoplamiento total de las perturbaciones, en muchas situaciones esto no es posible aun cuando las incertidumbres sean estructuradas, debido a las condiciones restrictivas (4.6). Entonces sólo pueden alcanzarse aproximaciones del desacoplamiento perfecto.

4.1.1.2 Desacoplamiento aproximado

Si las incertidumbres no son estructuradas o no pueden verificarse las condiciones (4.6), deben establecerse umbrales $\Lambda > 0$, y el problema de detección y aislamiento debe reformularse como sigue:

- $\| r(s) \| < \Lambda$ si $f(s) = 0$
 $\| r(s) \| \geq \Lambda$ si $f(s) \neq 0$
- $\| r_i(s) \| < \Lambda_i$ si $f_i(s) = 0$
 $\| r_i(s) \| \geq \Lambda_i$ si $f_i(s) \neq 0$

Ahora, el objetivo de diseño consiste en hallar una matriz de parametrización $Q(s)$ de modo de minimizar los umbrales frente a entradas desconocidas $d(s)$.

La generación de residuos robustos en el dominio de la frecuencia tiene algunas características atractivas. En primer lugar, ofrece métodos potentes para resolver el problema de la robustez utilizando la teoría de H_∞ . Por otra parte, permite tomar las especificaciones de frecuencia como criterios adicionales para una mejor discriminación entre fallas.

Para delinear el planteo básico de esta metodología de diseño, supongamos que se desea maximizar el siguiente índice de performance

$$J = \frac{\| \partial r / \partial f \|}{\| \partial r / \partial d \|}$$

donde $\| \cdot \|$ indica alguna norma de la sensibilidad con respecto a f y a d respectivamente. El objetivo de diseño es encontrar un generador de residuos tal que J sea máximo, lo que tiene sentido ya que se maximiza la sensibilidad a las fallas, mientras simultáneamente se minimiza la sensibilidad frente a perturbaciones. Utilizando (4.4), J puede expresarse

$$J = \frac{\| Q(s) \hat{M}_u(s) G_f(s) \|}{\| Q(s) \hat{M}_u(s) G_d(s) \|} \quad (4.7)$$

Si tomamos la norma \mathcal{L}_2 y reducimos la matriz $Q(s)$ a un vector $q(s)$, el problema de optimización se transforma en

$$J = \frac{\| q(s) G_2(s) \|_2}{\| q(s) G_1(s) \|_2} \xrightarrow{q(s)} \max \quad (4.8)$$

donde $G_1 = \hat{M}_u(s) G_d(s)$ y $G_2 = \hat{M}_u(s) G_f(s)$. La maximización resulta en

$$V(j\omega) [G_2(j\omega) G_1^T(-j\omega) - \lambda(\omega) G_1(j\omega) G_2^T(-j\omega)] = 0$$

o sea, un problema de autovalores y autovectores. La solución de este problema en el dominio de la frecuencia es

$$q_{opt}(s) = q_f(s) V(s), \quad J_{opt} = \sup_{\omega} (\lambda(j\omega))$$

Aquí, el autovector $V(s)$ es el selector para el generador de residuos óptimo y el máximo autovalor $\lambda(\omega_0)$ es el valor del índice de performance en la frecuencia óptima ω_0 en la cual se maximiza J . Por lo tanto, el filtro selecciona sólo aquella parte del espectro de frecuencia del residuo r que proporciona el mejor compromiso entre insensibilidad frente a perturbaciones y sensibilidad frente a las fallas.

Una técnica alternativa considera el siguiente problema de optimización

$$J = \frac{\| Q(s) G_2(s) \|_{\infty}}{\| Q(s) G_1(s) \|_{\infty}} \xrightarrow{Q(s)} \max \quad (4.9)$$

que ha sido resuelto utilizando la teoría H_{∞} por Ding [19].

La solución del problema (4.9) tiene también un importante significado físico. De hecho, la norma H_{∞} define una medida de la diferencia entre las matrices de transferencia $G_f(s)$ y $G_d(s)$. Para alcanzar la máxima diferencia, que se produce a una frecuencia ω_0 , se utiliza un pasabanda $Q_0(s)$. Como resultado, esta estrategia permite utilizar toda la información disponible en frecuencia para aumentar la capacidad del generador de residuos en cuanto a su robustez y distinguibilidad entre fallas.

4.2 Desacoplamiento de perturbaciones en sistemas no lineales

4.2.1 Desacoplamiento total

Sea un sistema no lineal de una entrada y una salida, de la forma:

$$\begin{aligned}\dot{x} &= F(x) + G(x)u + E(x)d \\ y &= h(x)\end{aligned}\tag{4.10}$$

donde d representa las perturbaciones o entradas desconocidas.

Consideremos el problema consistente en lograr que la salida y sea completamente independiente de las perturbaciones d . Este es equivalente, en DEF, al problema de obtener un residuo totalmente desacoplado de las perturbaciones. Por simplicidad, trabajaremos en esta sección con un modelo sin fallas.

Tal como se presenta en [33], debemos examinar bajo qué condiciones existe una ley de realimentación u , siendo:

$$u = \alpha(x) + \beta(x)v$$

tal que en el sistema a lazo cerrado la salida y se encuentre completamente desacoplada de la perturbación d .

Supongamos que el sistema tiene grado relativo r en x^0 , y que el campo vectorial $E(x)$ es tal que

$$L_E L_F^i h(x) = 0 \quad 0 \leq i \leq r - 1\tag{4.11}$$

$\forall x$ en un entorno de x^0 .

En la ecuación anterior, $L_F h(x)$ indica la derivada de Lie de la función $h(x)$ en la dirección del campo vectorial $F(x)$.

Nota 4.2.1 *Derivadas de Lie [16]*

Dadas las funciones $h : \mathbb{R}^n \rightarrow \mathbb{R}$ y $f : \mathbb{R}^n \rightarrow \mathbb{R}^n$, llamaremos *Derivada de Lie* de $h(x)$ en la dirección del campo vectorial $f(x)$ a la función $L_f h : \mathbb{R}^n \rightarrow \mathbb{R}$

$$L_f h(x) = \frac{\partial h(x)}{\partial x} \cdot f(x)$$

Como suponemos que todas las funciones intervinientes son infinitamente diferenciables, es posible aplicar repetidamente la operación anterior, para escribir en consecuencia

$$\begin{aligned} L_g L_f h(x) &= \frac{\partial L_f h(x)}{\partial x} \cdot g(x) \\ L_f^k h(x) &= \frac{\partial L_f^{k-1} h(x)}{\partial x} \cdot f(x) \end{aligned}$$

El cumplimiento de (4.11) garantiza que las entradas desconocidas o incertidumbres en el modelado que puedan aparecer en el sistema no se verán reflejadas en la salida. Si además se verifican las condiciones para la existencia de una forma normal definida globalmente [33], el desacoplamiento de perturbaciones puede ser global.

4.2.2 Atenuación de perturbaciones

El cumplimiento de una condición de la forma (4.11) impone restricciones severas al sistema, por lo cual sólo se verificará en casos especiales. Trataremos entonces aquí el problema menos exigente de hallar una ley de realimentación que minimice la influencia de las perturbaciones en la salida. Dicha ley debería garantizar también la estabilidad asintótica (global) del sistema a lazo cerrado correspondiente.

Para alcanzar nuestro objetivo de optimización es necesario establecer un criterio preciso para medir la influencia de una entrada dada (en este caso, la perturbación) en la salida del sistema. En los sistemas no lineales se utiliza con frecuencia el concepto de *ganancia* \mathcal{L}_2 , que definiremos a continuación.

Consideremos ahora un sistema de una entrada y una salida, descrito por las siguientes ecuaciones

$$\begin{aligned} \dot{x} &= F(x) + G(x)u \\ y &= h(x) \end{aligned} \tag{4.12}$$

donde F y G son campos vectoriales suaves y $h(x)$ es una función suave definida en \mathbb{R}^n . Supongamos que $F(0) = 0$ y $h(0) = 0$. Sea \mathcal{L}_2 el conjunto de todas las funciones constantes a tramos $u : [0, \infty] \rightarrow \mathbb{R}$ que satisfacen

$$\int_0^\infty u^2(s) ds < \infty$$

y $x(t, x^0, u(.))$ el valor alcanzado por el estado x a tiempo $t > 0$ bajo el efecto de la entrada $u(.) \in \mathcal{L}_2$, comenzando del estado inicial x^0 a $t = 0$.

Definición 4.2.1 *El sistema (4.12) tiene ganancia \mathcal{L}_2 menor o igual que γ si $\forall u \in \mathcal{L}_2$ la respuesta $x(t, 0, u(.))$ desde el estado inicial $x(0) = 0$ existe $\forall t \geq 0$ y verifica*

$$\int_0^t \|h(x(s, 0, u(.)))\|^2 ds \leq \gamma^2 \int_0^t \|u(s)\|^2 ds \quad \forall t > 0 \quad (4.13)$$

Enunciaremos a continuación la proposición demostrada en [33], que presenta una condición suficiente para que un sistema de la forma (4.12) tenga ganancia \mathcal{L}_2 menor o igual que γ .

Proposición 4.2.1 *Sea un sistema de la forma (4.12). Supongamos que existe una función suave $V(x)$ positiva definida y propia que satisface*

$$L_F V(x) + \frac{1}{4\gamma^2} [L_G V(x)]^2 + [h(x)]^2 < 0 \quad \forall x \neq 0 \quad (4.14)$$

Entonces, el sistema (4.12) tiene un equilibrio global y asintóticamente estable en $x = 0$ y tiene ganancia $\mathcal{L}_2 \leq \gamma$.

Demostración: Observemos que (4.14) es equivalente a

$$\frac{\partial V}{\partial x} (F(x) + G(x)u) + [h(x)]^2 - \gamma^2 u^2 < 0 \quad \forall x \neq 0, \forall u \in \mathbb{R} \quad (4.15)$$

De hecho, para cada x fijo el miembro izquierdo de la desigualdad anterior es un polinomio de segundo grado que tiene un máximo en

$$u(x) = u^*(x) = \frac{1}{2\gamma^2} L_G V(x)$$

Entonces (4.15) se verifica $\forall u$ si y sólo si el valor de su miembro izquierdo en $u = u^*$ es negativo, lo cual es precisamente la condición (4.14).

Elijamos una entrada $u^0(.) \in \mathcal{L}_2$, sea $x^0(.)$ la respuesta correspondiente del sistema (4.12) con estado inicial $x(0) = 0$ y supongamos que existe $x^0(t) \forall 0 \leq t < T$. Como (4.15) implica

$$\frac{dV(x^0(t))}{dt} \leq \gamma^2 [u^0(t)]^2 - [h(x^0(t))]^2 \quad (4.16)$$

Integrando con respecto a t en el intervalo $[0, t)$ con $t < T$,

$$V(x(t)) \leq \gamma^2 \int_0^t [u^0(s)]^2 ds - \int_0^t [h(x^0(s))]^2 ds$$

porque $V(0) = 0$.

Como $V(x(t))$ es positiva definida y propia, de (4.16) se sigue

$$\gamma^2 \int_0^t [u^0(s)]^2 ds \geq \int_0^t [h(x^0(s))]^2 ds + V(x(t)) \geq \int_0^t [h(x^0(s))]^2 ds,$$

tal como se pedía.

Por último, observemos que $L_F V(x) < 0 \quad \forall x \neq 0$ entonces $V(x)$ es una función de Lyapunov para el sistema autónomo $\dot{x} = F(x)$, que por lo tanto tiene un equilibrio global asintóticamente estable en $x = 0$.

$\diamond q.e.d.$

La existencia de una función $V(x)$ propia y definida positiva que satisfaga la ecuación (4.14), llamada *Desigualdad de Hamilton-Jacobi*, es una forma de determinar si un sistema dado es global y asintóticamente estable y tiene ganancia $\mathcal{L}_2 \leq \gamma$. Tomando la existencia de dicha función como un criterio para establecer una estimación de la influencia de la entrada en la salida de un sistema, es posible formular el problema de alcanzar un nivel determinado de atenuación de perturbaciones (en un sistema de la forma (4.12)), como el problema de hallar una realimentación $u = \alpha(x)$ tal que el correspondiente sistema de lazo cerrado

$$\begin{aligned} \dot{x} &= F(x) + G(x)\alpha(x) + E(x)d \\ y &= h(x) \end{aligned}$$

verifique para una función $V(x)$ propia, definida positiva, una desigualdad de la forma (4.14), es decir

$$L_f V(x) + \frac{1}{4\gamma^2} [L_E V(x)]^2 + [h(x)]^2 < 0 \quad \forall x \neq 0 \quad (4.17)$$

donde $f(x) = F(x) + G(x)\alpha(x)$.

4.2.2.1 Atenuación de perturbaciones para observadores no lineales

Como ya se dijo, la robustez frente a perturbaciones es una característica deseable para aquellos observadores no lineales que se utilicen como detectores de fallas, ya que permite una detección más confiable. Dado que las condiciones para lograr el desacoplamiento total son restrictivas, es interesante analizar la atenuación de la influencia de las perturbaciones en el error de observación.

Analizaremos entonces el problema “dual” del desarrollado en [33] para controladores, para lo cual utilizamos el modelo de error del sistema planta-observador, que puede representarse por las ecuaciones

$$\begin{aligned}\dot{e} &= \Gamma(e, x, \hat{x}, u) + E(x)d \\ r &= H(e)\end{aligned}\tag{4.18}$$

siendo $e = x - \hat{x}$, d las perturbaciones y $E(x)$ la matriz de distribución de las mismas. Nótese que la salida es el residuo, es decir el error de observación, y en consecuencia nuestro objetivo es atenuar la ganancia entre r y d .

En primer lugar y para establecer un criterio preciso que permita medir la influencia de las perturbaciones en el error de observación, diremos que el sistema descrito según (4.18) tiene ganancia \mathcal{L}_2 menor o igual que γ si para toda $d(\cdot) \in \mathcal{L}_2$ la respuesta $e(s, 0, d(\cdot))$ existe y satisface $\forall t > 0$

$$\int_0^t \|H(e(s, 0, d(\cdot)))\|^2 ds \leq \gamma^2 \int_0^t \|d(s)\|^2 ds$$

Una condición suficiente para que se cumpla lo anterior está dada por la existencia de una función $V(e)$, propia y definida positiva, que satisfaga una desigualdad de Hamilton-Jacobi. Estudiaremos el caso en que d no es escalar, para lo cual se debe analizar cómo cumplir la condición (4.17), que se requiere para tener estabilidad global y asintóticamente estable en el origen y ganancia \mathcal{L}_2 menor o igual que γ . Nuestro objetivo es, entonces, extender los resultados obtenidos en [33] a un campo multivariable.

La aplicación del teorema de Artstein-Sontag (Ver **Apéndice C**), basado en el concepto de función de control de Lyapunov [33], considerando que debe ser adaptado al problema de atenuación de perturbaciones con estabilidad y referido al modelo de error, permite enunciar la siguiente Proposición.

Proposición 4.2.2 *Sea un sistema de la forma (4.18). Supongamos que existe una función suave $V(e)$ positiva definida y propia que satisface*

$$L_{\Gamma}V(e) + \frac{1}{4\gamma^2}[L_EV(e)]^2 + [H(e)]^2 < 0 \quad \forall e \neq 0 \quad (4.19)$$

donde

$$L_{\Gamma}V(e) = \frac{\partial V(e)}{\partial e} \Gamma(e, x, \hat{x}, u) \quad (4.20)$$

$$L_EV(e) = \frac{\partial V(e)}{\partial e} E(x) \quad (4.21)$$

$$[H(e)]^2 = e^T e = \|e\|^2 \quad (4.22)$$

Entonces, el sistema (4.18) tiene un equilibrio global y asintóticamente estable en $e = 0$ y tiene ganancia $\mathcal{L}_2 \leq \gamma$.

Demostración: Tal como en la Proposición 4.2.1, la condición (4.19) es equivalente a

$$\frac{\partial V}{\partial e} [\Gamma(e, x, \hat{x}, u) + E(x)d] + \|e\|^2 - \gamma^2 \|d\|^2 < 0 \quad \forall d \quad (4.23)$$

Teniendo en cuenta que d no es escalar, el miembro izquierdo de la desigualdad anterior puede escribirse como una función de d de la forma

$$f(d) = C + B^T d - d^T A d < 0 \quad (4.24)$$

donde

$$A = \gamma^2 I \quad B^T = \frac{\partial V}{\partial e} E(x) = L_EV(e)$$

$$C = \|e\|^2 + L_{\Gamma}V(e)$$

Estudiaremos bajo qué condiciones se verifica esta desigualdad en forma global con respecto a las perturbaciones.

La condición (4.24) es equivalente a

$$-C - B^T d + d^T A d > 0 \quad (4.25)$$

y definiendo

$$\bar{f}(d) = -C - B^T d + d^T A d$$

el cumplimiento de (4.25) $\forall d$ equivale a requerir que el mínimo de \bar{f} sea positivo. Dicho mínimo se realiza para $d^* = \frac{A^{-1}B}{2}$ y reemplazando este valor en (4.25) se tiene

$$\bar{f}(d^*) = \frac{-B^T A^{-1} B}{4} - C > 0$$

Entonces, el máximo de (4.24) es negativo dado por

$$f(d^*) = -\bar{f}(d^*) = \frac{B^T A^{-1} B}{4} + C \quad (4.26)$$

con lo cual se verifica (4.23) $\forall d$. Reemplazando las expresiones de A,B, y C en (4.26) puede verse que ésta es, precisamente, la condición (4.19).

$\diamond q.e.d.$

Nota 4.2.2 Analogía para el caso lineal.

Si el sistema fuera

$$\begin{aligned} \dot{x} &= Ax + Bu \\ y &= Cx \end{aligned}$$

Una forma cuadrática, definida positiva $V(x) = x^T P x$ satisface la desigualdad (4.16) si y sólo si existe P simétrica, definida positiva de modo que:

$$PA + A^T P + C^T C + \frac{1}{\gamma^2} P B B^T P < 0$$

Esta es una condición necesaria y suficiente conocida, para que un sistema sea asintóticamente estable y tenga ganancia \mathcal{L}_2 estrictamente menor que γ (Bounded Real Lemma).

Capítulo 5

Sistemas de tratamiento de efluentes

En este capítulo presentaremos las características generales de los procesos de tratamiento biológico de aguas residuales, y realizaremos un breve resumen de los principales métodos de pretratamiento y tratamiento primario. Luego analizaremos el tratamiento secundario, haciendo énfasis en el sistema de lodos activados. Estudiaremos este último con mayor detalle, con el objetivo de presentar los balances de materia y ecuaciones cinéticas que permitirán desarrollar el modelo matemático de la planta que se muestra en la Sección 5.4. Por último, describiremos algunas fallas típicas que suelen presentarse en el proceso, a fin de completar un modelo de fallas que permita posteriormente instrumentar un sistema de detección y aislamiento de las mismas.

5.1 Los tratamientos biológicos. Generalidades.

El tratamiento biológico de residuos líquidos se aplica a la remoción de compuestos biodegradables y se utiliza con frecuencia en todo el mundo para estabilizar desagües domésticos, como asimismo residuos orgánicos de varios tipos de industrias [49], [44].

La biodegradación o descomposición orgánica no es más que el resultado de los procesos de digestión, asimilación y metabolización del compuesto orgánico por microorganismos tales como bacterias, hongos, protozoos y otros. Para que tal utilización del compuesto químico como alimento sea

posible, es necesario que se cumplan dos condiciones básicas:

- que sus moléculas posean elevado contenido de energía potencial disponible, como es el caso de todas las moléculas formadas por extensas cadenas carbónicas, denominadas *orgánicas*.
- que el microorganismo disponga de enzimas capaces de catalizar la reacción química que lleva al rompimiento de las cadenas orgánicas y a la liberación de su energía potencial.

Esta liberación de energía potencial se logra merced a la oxidación biológica o a la respiración, por lo cual esta última constituye una de las etapas fundamentales en cualquier proceso de biodegradación.

La respiración aeróbica, que será de nuestro interés en el proceso de lodos activados, resulta necesariamente en una demanda de oxígeno del ambiente. Si el número de organismos es muy grande en relación con el espacio y si la renovación del oxígeno en este espacio es dificultosa, la continuación del proceso aeróbico puede verse limitada. En el caso del agua es necesario tener en cuenta, además, la baja solubilidad del gas y la reducida velocidad de su difusión en dicho ambiente. Por este motivo, cuando la población de microorganismos aeróbicos es muy concentrada disminuye en alto grado la concentración de oxígeno, llegando a agotarse completamente cuando el medio se vuelve anaeróbico.

Sin embargo, la población de microorganismos en un ambiente dado es proporcional a la cantidad de alimento orgánico, biodegradable, en el mismo. Por ello, se puede decir que si se introduce una cierta cantidad de materia biodegradable en el agua, la *demanda bioquímica de oxígeno* (DBO) es proporcional. Este constituye uno de los parámetros más importantes para medir el posible consumo de oxígeno provocado por una carga orgánica, o sea por determinada cantidad de desecho biodegradable. Al mismo tiempo, permite evaluar el efecto genérico que esta carga de polución pudiera provocar sobre los ecosistemas acuáticos.

En cuanto al tratamiento biológico, esencialmente se trata de un proceso de descomposición acelerada, que consiste en ofrecer a los microorganismos responsables de la descomposición todos los elementos y las condiciones que normalmente constituyan factores limitantes del proceso. La descomposición puede tener lugar en un medio aeróbico o anaeróbico. Esta última

modalidad suele reservarse para el tratamiento de lodos o de residuos semilíquidos, en los cuales es difícil garantizar el acceso de oxígeno al interior de la masa pastosa de materia orgánica.

Para el tratamiento se requieren desde simples lagunas o zanjas excavadas en el suelo, hasta los más complejos biorreactores o quimiostatos, con diferentes niveles de mecanización y automatización.

Todos los métodos de tratamiento implican la alimentación de biomasa con los compuestos orgánicos existentes en el desecho, los cuales se comportan como sustrato de la actividad microbiana. Parte de este sustrato consumido es metabolizado y transformado en energía y productos minerales mediante la respiración; otra parte es acumulada como incremento de biomasa. En los procesos aeróbicos, como ya se dijo, el principal factor limitante es el oxígeno, el cual debe ser suministrado para activar y acelerar el mecanismo de estabilización. Las variadas modalidades tecnológicas de tratamiento aeróbico difieren esencialmente en cuanto a la forma de proporcionar este oxígeno. Se destacan las siguientes alternativas: por simple contacto con el aire atmosférico (lechos de contacto); por turbulencia o introducción de aire comprimido u oxígeno puro (lodos activados); por fotosíntesis de algas (lagunas de estabilización). En algunos casos, otros elementos pueden volverse limitantes cuando el oxígeno es suministrado en cantidades suficientes; los principales son el nitrógeno y el fósforo.

En el proceso de descomposición la biomasa activa está constituida principalmente por microorganismos saprófitos (bacterias y hongos), además de depredadores de estos microorganismos. Esta biomasa puede permanecer dispersa en el medio líquido, como ocurre en las *lagunas de estabilización*, quedar fija a un soporte físico, como en los *lechos de contacto* o *lechos bacterianos*, sobre los cuales se escurre una delgada película de agua residual, o formar masas granuladas o flóculos en suspensión, como en el sistema de *lodos activos*.

En cualquiera de estos procesos, la materia orgánica poluyente presente en el efluente es removida en parte por oxidación biológica (o respiración de los microorganismos), transformándose en dióxido de carbono, agua y energía, y en parte por transformación en biomasa, la cual es continuamente eliminada junto con el efluente (en las lagunas de estabilización) o es sedimentada constituyendo un lodo que se extrae mecánicamente (en el proceso de lodos activados y algunos tipos de lechos de contacto).

La secuencia básica observada en una planta de tratamiento convencional de lodos activados es la siguiente: el efluente, después de introducido en la planta, pasa primero por un sistema de cribas o rejillas, con el fin de remover objetos de mayores dimensiones (papeles, hojas, etc.); luego, es conducido a una sedimentadora de arena. La etapa siguiente es la decantación primaria en grandes estanques, donde se deposita el lodo primario, constituido por materias orgánicas insolubles (sólidos sedimentados) asociados a arcillas y otros materiales finos. Este lodo es llevado a los digestores, donde es sometido a tratamiento anaeróbico, completándose así la etapa primaria.

El tratamiento secundario es entonces aplicado a un líquido constituido por agua con materias en solución, orgánicas e inorgánicas, además de finas partículas en suspensión (no sedimentables). La materia orgánica presente en esta forma (soluble y en suspensión permanente) representa alrededor del 60 % de la carga de DBO del desagüe. Este líquido es sometido a algún proceso aeróbico; de entre todos los posibles describiremos aquí el de lodos activados, que es el más complejo pero también el más utilizado en las plantas actuales.

En las plantas más recientes, suele observarse una etapa terciaria, donde se produce una reducción de la carga microbiana, en especial de los patógenos, por medio de una desinfección o filtración por membranas.

5.2 Principales métodos de pretratamiento y tratamiento primario de aguas residuales

Los pretratamientos de efluentes consisten en la reducción de sólidos en suspensión o el acondicionamiento de las aguas residuales para su descarga en los receptores o bien para pasar a un tratamiento secundario a través de una neutralización u homogeneización. Los tipos fundamentales de tratamientos primarios que trataremos aquí son: cribado, sedimentación, flotación, neutralización y homogeneización [49].

5.2.1 Cribado

El cribado, también llamado desbrozo, se emplea para la reducción de sólidos en suspensión de distintos tamaños. La distancia o aberturas de las rejillas dependen de su función, y su limpieza se realiza en forma manual o mecánica. Los productos recogidos se destruyen por incineración, se tratan por procesos de digestión anaerobia o se dirigen directamente a un vertedero. Las materias sólidas recogidas suelen clasificarse en finos y gruesos.

Mediante las rejillas de finos pueden eliminarse entre un 5 y un 25 % de sólidos en suspensión, en tanto que un 40 a un 60 % se extraen por sedimentación. Por este motivo, y también porque el atascamiento suele ser un problema, el uso de tamices finos o con abertura pequeña no es muy común.

Las rejillas o cribas de gruesos se utilizan como elementos de protección para evitar que sólidos de grandes dimensiones dañen las bombas y otros equipos mecánicos. A veces se reemplazan las rejillas de gruesos por trituradoras, que rompen o desgarran los sólidos en suspensión para luego ser eliminados por sedimentación.

5.2.2 Sedimentación

La sedimentación, cuyo principio se basa en la diferencia de peso específico entre las partículas sólidas y el líquido en que se encuentran, es utilizada en este caso para separar sólidos en suspensión de las aguas residuales. Este proceso puede producirse en una o varias etapas o en varios de los puntos del proceso de tratamiento. En una planta típica de lodos activos, la sedimentación se utiliza en tres de las fases del tratamiento:

- en los desarenadores, en los cuales la materia inorgánica (a veces, arena) se elimina del agua residual.
- en los clarificadores o sedimentadores primarios, que preceden al reactor biológico y en los cuales los sólidos (orgánicos y otros) se separan.
- en los clarificadores o sedimentadores secundarios, que siguen al reactor biológico, en los cuales los lodos se separan del efluente tratado.

5.2.3 Flotación

La flotación es un proceso para separar sólidos de baja densidad o partículas líquidas de una fase líquida. La separación se lleva a cabo introduciendo un gas (normalmente aire) en la fase líquida, en forma de burbujas. La fase líquida se somete a un proceso de presurización, en presencia de suficiente aire para conseguir la saturación en aire del agua. Luego, este líquido saturado de aire se despresuriza hasta la presión atmosférica. En esta situación y debido a la despresurización, se forman pequeñas burbujas de aire que se desprenden de la solución. Los sólidos en suspensión o las partículas líquidas (como por ejemplo, aceites o petróleo) flotan, debido a que estas pequeñas burbujas, asociándose a los mismos, les obligan a elevarse hacia la superficie. Los sólidos en suspensión concentrados pueden separarse de la superficie por sistemas mecánicos. El líquido clarificado puede separarse cerca del fondo, y parte del mismo puede reciclarse.

En el campo de los tratamientos de aguas residuales, la flotación se usa para los siguientes objetivos:

1. separación de grasas, aceites, fibras, y otros sólidos de baja densidad, de las aguas residuales;
2. espesado de los lodos procedentes de los procesos de lodos activos;
3. espesamiento de los lodos floculados químicamente resultantes de los tratamientos de coagulación química.

Una superior calidad de los efluentes (efluentes con menores porcentajes de sólidos en suspensión) y una economía de energía pueden conseguirse con sistemas de flotación con reciclaje.

5.2.4 Neutralización y homogeneización

El tratamiento de neutralización se utiliza normalmente en los siguientes casos que se presentan en la depuración de aguas residuales:

1. Antes de la descarga de aguas residuales en un medio receptor. La justificación para la neutralización es que la vida acuática es muy sensible a variaciones de pH fuera de un intervalo cercano a $pH = 7$.

2. Antes de la descarga de aguas residuales industriales al alcantarillado municipal. La especificación de pH de las descargas industriales en las alcantarillas se hace de forma frecuente.
3. Antes del tratamiento químico o biológico. Para los tratamientos biológicos, el pH del sistema se mantiene en un intervalo comprendido entre 6,5 y 8,5 para asegurar una actividad biológica óptima. El proceso biológico en sí mismo puede conseguir una neutralización, y en cualquier caso tiene una capacidad reguladora como resultado de la producción de dióxido de carbono, que da lugar a la formación de carbonatos y bicarbonatos en la solución. El grado de preneutralización requerido para el tratamiento biológico depende de dos factores: la alcalinidad o acidez presente en el agua residual, y los mg/l de DBO que deben eliminarse en el tratamiento biológico.

En cuanto a los métodos para neutralización de aguas residuales, se utilizan los siguientes:

- **homogeneización:** Consiste en mezclar las corrientes disponibles en la planta, algunas de las cuales son ácidas y otras alcalinas. Suele utilizarse, además, como método para conseguir que la corriente de alimentación a la planta tenga un caudal y una DBO relativamente constantes, con lo cual se disminuye la influencia de las fluctuaciones de dichos parámetros, que representan perturbaciones en el proceso.
- **métodos de control directo de pH:** Consisten en la adición de ácidos (o bases) para neutralizar las corrientes alcalinas o ácidas. Los más utilizados son: lechos de caliza, neutralización con cal, neutralización con sosa cáustica, neutralización con carbonato de sodio y neutralización con amoníaco. Esta última es contaminante, y por lo tanto su uso es limitado.

5.3 Tratamiento secundario: el proceso de lodos activos

El proceso se inicia en un tanque de aireación o biorreactor, donde el efluente industrial líquido es sometido a una intensa oxigenación, ya sea por medio de tubos cribados que distribuyen aire comprimido en el fondo del estanque

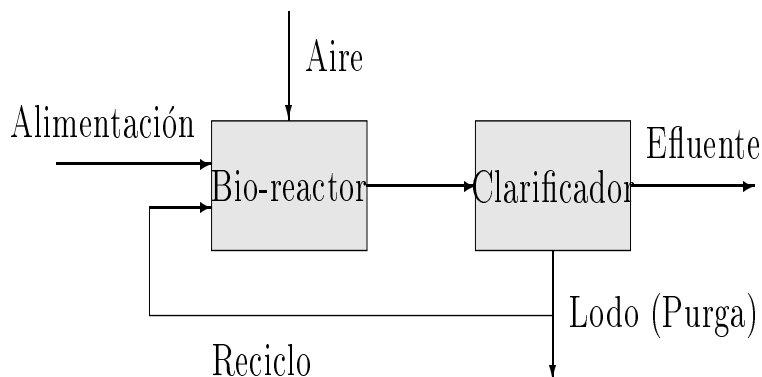


Figura 5.1: Diagrama de la planta de tratamiento de aguas residuales.

o mediante rotores superficiales (hélices o rotores de eje vertical o cepillos cilíndricos de eje horizontal). Después de un período variable de aireación, el efluente de este tanque pasa a un decantador o clarificador, donde tiene lugar la sedimentación del lodo formado durante la aireación, constituido esencialmente por biomasa de bacterias y otros microorganismos. Parte de la biomasa sedimentada es reciclada y vuelve al biorreactor, donde servirá de “semilla” para acelerar la formación de nueva biomasa activa. En el clarificador se remueve el lodo por sedimentación (corriente de purga), en tanto que el sobrenadante constituye el efluente tratado o estabilizado (Ver Figura 5.1).

Los principales procesos biológicos, bioquímicos y fisicoquímicos que intervienen en el proceso de tratamiento tienen lugar, pues, en el tanque de aireación. Este no es más que un reactor aeróbico comparable a un *quimiostato*, donde el sustrato está constituido por la materia orgánica del agua residual, de naturaleza compleja, y la acción enzimática es ejercida por una población mixta de microorganismos descomponedores y consumidores de varios niveles. El proceso adquiere mayor complejidad en la medida en que la competencia entre dichos microorganismos no está determinada solamente por la naturaleza física, soluble o no soluble del sustrato, sino también por sus características químicas. Así, por ejemplo, el predominio de determinadas bacterias suele ser el resultado de la competencia entre ellas en busca de fuentes de nitrógeno o fósforo.

En el caso de un biorreactor “real” alimentado de manera continua con una concentración de sustrato teóricamente constante, el equilibrio en cuan-

to a la composición química y microbiológica que interviene en el proceso de depuración está determinado por la optimización de una relación costo-beneficio. Esta consiste en obtener en el menor tiempo posible un efluente de calidad compatible con la calidad que se desea mantener en la masa de agua que lo recibe. En consecuencia, la composición biológica característica de esta situación de equilibrio puede servir como índice de la eficiencia o desempeño del sistema en términos prácticos. En general, se considera indicio de buen desempeño la presencia de flóculos compactos de bacterias.

La prolongación del proceso de estabilización, es decir de la aireación sin abastecimiento de nuevo sustrato, conduce a una situación llamada de *respiración endógena*, caracterizada por la autodigestión de los flóculos. Si bien este fenómeno no está completamente definido desde el punto de vista biológico, representa un conjunto de procesos que contribuyen a reducir la biomasa.

En el análisis de la dinámica del tratamiento biológico de efluentes deben considerarse los aspectos energéticos, bioquímicos y fisicoquímicos de la floculación.

- Aspectos energéticos

Las relaciones entre crecimiento de biomasa y disipación de energía en los procesos de oxidación biológica se rigen por el segundo principio de la termodinámica. En este caso, en el cual el sustrato está constituido por un residuo complejo y la población de descomponedores y consumidores es mixta, no solamente los fenómenos de inhibición y competencia han de elevar el consumo de energía de conservación, sino también en cada eslabón de la cadena alimentaria, que se forma por la presencia de depredadores, aumenta la entropía correspondiente. Por esto deben evaluarse cuidadosamente los porcentajes de conversión de sustrato en biomasa. Al parecer, la cantidad de sólidos biológicos generada en estos sistemas incluye una porción significativa de materia orgánica no metabolizada del residuo, que es adsorbida y precipitada conjuntamente con los flóculos bacterianos formados.

Por otra parte, una de las grandes dificultades para la evaluación del desempeño del sistema la constituye la medición de la *biomasa activa* durante el proceso de tratamiento, la cual es evaluada en forma muy aproximada en términos de *sólidos suspendidos volátiles* en la

cámara de aireación. Esto presupone que todo sustrato orgánico se encuentra en forma soluble, lo que en general no sucede. Hay que considerar finalmente una cantidad de residuos de células y microorganismos muertos que, obviamente, no forman parte de la biomasa activa.

- Aspectos bioquímicos

La materia orgánica se estabiliza mediante un proceso de tratamiento biológico por la acción específica de enzimas producidas por los microorganismos participantes sobre el sustrato complejo representado por el conjunto de compuestos biodegradables que componen el efluente. No se trata, entonces, de una sencilla relación *enzima-sustrato* sujeta a las leyes cinéticas que rigen algunas reacciones conocidas en la aplicación industrial.

El rendimiento de un proceso enzimático depende en general de la concentración del sustrato y de las enzimas, de la presencia de sustancias activadoras o inhibidoras y de otros factores, como el pH y la temperatura. Cuando hay exceso de sustrato, el rendimiento es función de la concentración de enzimas en el medio, en tanto que a una concentración fija de enzimas, el mismo depende de la concentración de sustrato. La velocidad de reacción alcanza un máximo correspondiente a la saturación de la enzima por el sustrato (fórmula de Michaelis-Menten). Sin embargo, para aplicar estas relaciones al proceso de lodos activados deben tenerse en cuenta las restricciones que introduce la naturaleza compleja del residuo orgánico en tratamiento, así como también la presencia de poblaciones mixtas de microorganismos. Además, puede detectarse la presencia de inhibidores enzimáticos, tales como metales pesados, que aparecen en concentraciones variables y de manera permanente o esporádica.

Otro aspecto que merece atención en la bioquímica del proceso de tratamiento es la *nitrificación*, es decir la oxidación biológica del nitrógeno amoniacal, que a su vez proviene de la descomposición de compuestos nitrogenados más complejos que existen en los desagües. El interés por obtener un mayor o menor grado de nitrificación depende del interés local y principalmente de las características del caudal receptor. Como ya se ha señalado, el sistema de tratamiento tiene por objeto

obtener un grado de depuración que permita conseguir un efluente de calidad compatible a la que se desea mantener en la masa de agua receptora. Ahora bien, el vertimiento de un efluente de alta concentración de nitrógeno amoniacal constituye una fuente de demanda de oxígeno para la transformación del amoníaco en nitritos y nitratos. Siendo así, si el caudal receptor es suficientemente voluminoso como para soportar esta demanda sin reducción significativa de su contenido de oxígeno, la nitrificación podrá tener lugar en el mismo río o lago; si por el contrario la relación de dilución es pequeña, la nitrificación deberá conseguirse durante el proceso de tratamiento. Finalmente, el nitrógeno en su forma nítrica es mucho menos tóxico para la fauna acuática que en la forma amoniacal.

- Aspectos fisicoquímicos y sistemáticos de la floculación

La coagulación, floculación y sedimentación de la biomasa en un sistema de lodos activados constituyen los aspectos más importantes en el proceso de remoción de materias poluentes pues, como ya se ha dicho, parte de la materia biodegradable del efluente se convierte en biomasa. El proceso de la biofloculación ha sido extensamente investigado, considerando la naturaleza, el aspecto microscópico y la composición microbiológica de los flóculos en relación a su eficiencia en la remoción de DBO. Además, debe estudiarse la capacidad de sedimentación de los flóculos, o sea su capacidad de precipitarse durante la decantación secundaria, originando el lodo biológico llamado lodo activado.

La coagulación es de gran importancia para la sedimentación de la masa biológica y la separación del efluente líquido. La aireación del efluente lo oxida rápidamente debido a la intensa proliferación de bacterias aeróbicas que consumen materia nutritiva y entran en fase endógena por sobrepoblación. En esta fase tiene lugar la coagulación, que hace posible la precipitación de las partículas restantes en suspensión. En tales condiciones, pueden darse varias situaciones: la aireación deficiente mantendrá un exceso de materias nutritivas y, por lo tanto, exceso de valor energético, lo que impedirá la buena coagulación; la aireación excesiva provoca intensa coagulación, pero da origen a flóculos de baja capacidad purificadora en virtud de un muy reducido metabolismo de las bacterias que lo forman.

En general, los flóculos están compuestos por una gran variedad de especies microbianas, predominando las bacterias, microorganismos filamentosos (bacterias u hongos), amebas, ciliados. Esta composición varía con el grado de estabilidad bioquímica del sistema como asimismo con la composición química del residuo. Así, cuando el efluente es sometido a aireación prolongada, sin renovarse la carga, se observa una sucesión de microorganismos.

La forma y estructura del flóculo pueden variar mucho. Cuando predominan bacterias no filamentosas, el flóculo es compacto, de bordes lisos. Con frecuencia, se observan flóculos compactos de bordes irregulares muy recortados o dendríticos, que en virtud de su mayor superficie relativa suelen ser más eficientes para la adsorción de partículas y la clarificación del líquido residual. Por último, los flóculos de estructura filamentosa, en los cuales predominan bacterias u hongos en forma de hilos enredados, son muy voluminosos y de baja densidad. Si bien son los más eficientes para la remoción de materias coloidales y partículas en suspensión, no sedimentan con facilidad, dando origen a un lodo leve, intumesciente, que ocupa gran volumen en el decantador y es arrastrado junto con el efluente. El predominio de microorganismos filamentosos origina, entonces, el problema de *intumescencia del lodo*, cuya causa se encuentra en varios factores aún no perfectamente conocidos. Sin embargo, una pequeña proporción de microorganismos filamentosos en el biorreactor puede ser benéfica para el tratamiento, ya que los filamentos pueden constituir una base estructural sobre la cual se fijan otras bacterias, originando la forma dendrítica de los flóculos compactos de mayor eficiencia.

5.4 Modelo matemático de la planta

Las reacciones involucradas en la remoción biológica de nutrientes de un efluente pueden ser clasificadas en: remoción de carbono orgánico, nitrógeno y fósforo. Estudiaremos a continuación la primera de ellas, para cuyo análisis se utilizarán balances de materia y relaciones cinéticas que involucran a cuatro componentes: carbono soluble (sustrato), oxígeno, agua y biomasa (heterotrófica).

Las hipótesis para construir el modelo son las siguientes [44]:

- El biorreactor es un tanque de aireación de volumen constante, perfectamente mezclado. Esto último implica una concentración homogénea en el tanque e igual a la de la corriente de salida.
- El clarificador es un sedimentador perfecto, sin dinámica. En él no se consumen microorganismos.
- No hay microorganismos en el efluente.
- No hay oxígeno en el reciclo ni en la corriente de purga.
- El efluente, el reciclo y la purga tienen la misma concentración de sustrato.
- Existe una relación lineal entre el parámetro de transferencia másica del oxígeno y el caudal de aire [21].

La descripción de las variables a utilizar se encuentra en la Tabla 5.1.

A continuación, desarrollaremos los balances de materia utilizados en la construcción del modelo.

5.4.1 Balance de masa para el oxígeno en el biorreactor

$$\frac{d}{dt}(V S_O) = q_F S_{OF} - (q_F + q_R) S_O - r_O V + K_L a (S_{O,sat} - S_O) \quad (5.1)$$

Las variables utilizadas, así como los restantes parámetros, se definieron en la Tabla 5.1.

5.4.1.1 La transferencia de oxígeno:

La transferencia de masa es el movimiento de un componente entre dos fases. En los recipientes aireados de lodos activados, el ejemplo más destacable es el de la transferencia de oxígeno desde el aire hacia el agua, con el objeto de ser utilizado por la biomasa. En principio, el oxígeno debe desplazarse desde la masa de aire hacia la superficie del agua, disolverse en ésta y luego moverse desde la superficie hacia la masa de líquido. En tratamiento de aguas se asume en general que, como consecuencia de la turbulencia del aire

Tabla 5.1: Descripción de variables para el modelo de la planta

SÍMBOLO	DESCRIPCION
q_F	Caudal de alimentación [Ml/día]
S_{SF}	Conc. sustrato alimentación [mg/l]
S_{OF}	Conc. O_2 en alimentación [mg/l]
X_{HF}	Conc. biomasa alimentación [mg/l]
S_S	Conc. sustrato en tanque [mg/l]
S_O	Conc. de O_2 en el tanque [mg/l]
X_H	Conc. biomasa en tanque [mg/l]
q_A	Caudal de aire [Ml/día]
q_R	Caudal de reciclo [Ml/día]
q_W	Caudal de purga [Ml/día]
V	Volumen del tanque [Ml]
μ_H	Tasa máx. crecim. biomasa [1/día]
K_S	Cte. saturación sustrato [mg/l]
K_{OH}	Cte. saturación de O_2 [mg/l]
Y_H	Coefficiente de rendimiento
b_H	Tasa decaimiento biomasa [1/día]
f_p	Fracción de inertes
$S_{O,sat}$	Conc. saturación de O_2 [mg/l]
K_a	Parám. transf. de masa [1/Ml]
r_0	Veloc. consumo O_2 [mg/ l.día]
$K_L a$	Coef. transf. masa [1/día]

y las burbujas, el agua en la superficie de las mismas se encuentra saturada con oxígeno. Se puede entonces representar la transferencia de masa con la siguiente ecuación

$$\begin{pmatrix} \text{veloc. de} \\ \text{transf. de } O_2 \end{pmatrix} = \begin{pmatrix} \text{coef. de} \\ \text{transf. de masa} \end{pmatrix} \begin{pmatrix} \text{area superf.} \\ \text{aire - agua} \end{pmatrix} \begin{pmatrix} \text{dif. de} \\ \text{concentraciones} \end{pmatrix}$$

En símbolos matemáticos:

$$r_a = K_L a (S_{O,sat} - S_O)$$

En nuestro caso y de acuerdo a las hipótesis, supondremos que el coeficiente de transferencia de masa varía linealmente con el caudal de aire, siendo:

$$K_L a = K_a q_A$$

K_a es una constante que debe ser determinada para una instalación particular en un punto definido de operación.

El parámetro $K_L a$ se utiliza en varios fenómenos de transferencia de masa. K_L puede verse como un coeficiente de absorción y a es una relación área-volumen. En el caso del biorreactor, el área de contacto aire-agua está relacionada con el tamaño de las burbujas, que es una función del caudal de aire y del equipamiento de aireación. Pequeñas burbujas se destruyen más lentamente que las grandes, y tienen un tiempo de contacto más largo. También proporcionan una mayor área de contacto. Por otra parte, la velocidad de transferencia de oxígeno es diferente en el efluente que en el agua limpia; esta diferencia se expresa afectando al parámetro de transferencia con un coeficiente de proporcionalidad. Análogamente, la concentración de saturación para el oxígeno disuelto es diferente en el efluente que en el agua limpia.

5.4.1.2 Objetivo de control

Una variable clave en la operación del lodo activado es la concentración de oxígeno disuelto. Su manipulación se realiza actuando sobre el caudal de aire y entonces se elige a éste como variable de control. Cabe destacar que existe un incentivo económico para minimizar el consumo de aire, pues su

compresión aumenta notablemente el gasto de energía y éste influye significativamente sobre los costos operativos.

En cuanto a la elección del valor deseado de la concentración de oxígeno disuelto, ésta debe ser suficientemente elevada como para que el crecimiento de microorganismos autótrofos y heterótrofos no se vea limitado por ella. Sin embargo, un exceso de aireación y de mezclado pueden impedir la floculación, afectando el rendimiento del sistema.

En resumen, el objetivo del control de caudal de aire debe ser mantener la concentración de oxígeno a un nivel lo suficientemente elevado como para lograr el crecimiento de los organismos adecuados, y lo suficientemente bajo como para ahorrar energía y evitar un exceso de mezclado.

5.4.2 Balance de masa para el nutriente carbono (sustrato)

$$\frac{d}{dt}(V S_S) = q_F (S_{SF} - S_S) + r_S V \quad (5.2)$$

donde r_S es la velocidad de generación del nutriente, y las restantes variables se definieron en la Tabla 5.1.

5.4.3 Balance de masa para los microorganismos en el clarificador

$$(q_F + q_R) X_H = (q_R + q_W) X_{HR} \quad (5.3)$$

$$X_{HR} = \frac{q_F + q_R}{q_W + q_R} X_H \quad (5.4)$$

donde X_{HR} es la concentración de microorganismos en el reciclo.

5.4.4 Balance de masa para los microorganismos en el biorreactor

$$\frac{d}{dt}(V X_H) = q_F X_{HF} + q_R X_{HR} - (q_F + q_R) X_H + r_H V \quad (5.5)$$

donde r_H es la velocidad de crecimiento de los organismos heterótrofos.

Tabla 5.2: Expresiones cinéticas

PROCESO	COMPONENTES			CINETICA
	Sustrato	Oxígeno	Biomasa	
Crecimiento aeróbico heterotrófico	$\frac{-1}{Y_H}$	$\frac{Y_H-1}{Y_H}$	1	μ
Decaimiento heterotrófico	$1 - f_p$	-	-1	$b_H X_H$

$$\text{Donde } \mu = \mu_H \left(\frac{S_S}{K_S + S_S} \right) \left(\frac{S_O}{K_{OH} + S_O} \right) X_H$$

5.4.5 Ecuaciones cinéticas

Para completar el modelo, deben definirse las velocidades de reacción o de crecimiento. Dada la complejidad del proceso, éstas no pueden expresarse con las ecuaciones cinéticas clásicas que suponen, entre otros factores, el conocimiento preciso de la estequiometría de la reacción. Teniendo en cuenta los aspectos bioquímicos analizados en la Sección 5.3, puede concluirse que la ecuación de Michaelis-Menten proporciona una representación general de la cinética del consumo de sustrato. Para la velocidad de crecimiento de biomasa se desarrolla una expresión equivalente, conocida como ecuación de Monod, en la cual un factor de decaimiento indica la biomasa perdida en la etapa de respiración endógena. Todo esto se sintetiza convencionalmente en la literatura en forma tabular, tal como se presenta en la Tabla 5.2 [44].

La expresión para la cinética o crecimiento de un componente en particular (r_H , r_O , r_S) se obtiene sumando todos los términos en la columna apropiada luego de multiplicar cada uno por la expresión cinética de la última columna.

5.4.6 Ecuaciones del modelo

Con las expresiones cinéticas y balances de masa desarrollados previamente, pueden completarse las ecuaciones del modelo, que quedan

$$\begin{aligned} \frac{dS_O}{dt} = & \frac{q_F}{V} S_{OF} - \frac{(q_F + q_R)}{V} S_O + \\ & + \frac{Y_H - 1}{Y_H} \mu_H \left(\frac{S_S}{K_S + S_S} \right) \left(\frac{S_O}{K_{OH} + S_O} \right) X_H + K_a q_A (S_{O,sat} - S_O) \end{aligned}$$

$$\begin{aligned}
\frac{dS_S}{dt} &= \frac{q_F}{V} (S_{SF} - S_S) - \frac{\mu_H}{Y_H} \left(\frac{S_S}{K_S + S_S} \right) \left(\frac{S_O}{K_{OH} + S_O} \right) X_H + \\
&\quad + (1 - f_p) b_H X_H \\
\frac{dX_H}{dt} &= \frac{q_F}{V} X_{HF} - \frac{q_W}{V} \left(\frac{q_F + q_R}{q_W + q_R} \right) X_H + \\
&\quad + \mu_H \left(\frac{S_S}{K_S + S_S} \right) \left(\frac{S_O}{K_{OH} + S_O} \right) X_H - b_H X_H \quad (5.6)
\end{aligned}$$

El modelo matemático de la planta tendrá como variables de estado las concentraciones de oxígeno, sustrato y biomasa, que llamaremos respectivamente x_1 , x_2 , x_3 y como variable de control al caudal de aire, que llamaremos u .

De los restantes parámetros, el volumen es de diseño, la concentración de saturación del oxígeno es conocida para las condiciones dadas, y los demás son desconocidos. Realizando estudios de sensibilidad dinámica y estática, se puede decidir cuáles de ellos deben ser ajustados a las condiciones de operación y cuáles -los menos sensibles- pueden ser tomados directamente de la literatura.

Se obtiene el siguiente sistema de ecuaciones diferenciales [44]

$$\begin{aligned}
\dot{x}_1 &= \frac{q_F}{V} (S_{OF} - x_1) - \frac{q_R}{V} x_1 + \frac{Y_H - 1}{Y_H} \bar{\mu} + \\
&\quad + K_a q_A (S_{O,sat} - x_1) \\
\dot{x}_2 &= \frac{q_F}{V} (S_{SF} - x_2) - \frac{1}{Y_H} \bar{\mu} + (1 - f_p) b_H x_3 \\
\dot{x}_3 &= \frac{q_F}{V} X_{HF} - \frac{q_W}{V} \left(\frac{q_F + q_R}{q_W + q_R} \right) x_3 + \bar{\mu} - b_H x_3
\end{aligned}$$

donde x_1 es la concentración de oxígeno, x_2 es la concentración de sustrato, x_3 es la concentración de biomasa y

$$\bar{\mu} = \mu_H \left(\frac{x_2}{K_S + x_2} \right) \left(\frac{x_1}{K_{OH} + x_1} \right) x_3$$

El significado de los demás parámetros se encuentra en la Tabla 5.1.

5.4.6.1 La ecuación de salida

El problema de la medición en bioprocesos radica en que las variables más importantes no pueden determinarse, en general, por la instrumentación en línea convencional. En particular, en el sistema de lodos activados las concentraciones de sustrato y biomasa sólo pueden conocerse por medio de mediciones infrecuentes (análisis de laboratorio de una muestra de proceso), por lo cual estos valores no están disponibles instantáneamente a los efectos del control o la detección. Por su parte, la medición de oxígeno disuelto es una medición *in situ*. El dispositivo, en contacto con el agua, generalmente está protegido por una membrana y tiene un tiempo de respuesta muy corto (la medición es virtualmente instantánea). En consecuencia, se elige a la concentración de oxígeno disuelto como variable de salida, y la ecuación correspondiente es

$$y = x_1$$

5.4.6.2 Variable de control

Como se dijo en la Sección 5.4.1.2, debido a la importancia de la concentración de oxígeno disuelto, cuya manipulación se realiza actuando sobre el caudal de aire, se elige a éste último como variable de control. Entonces, el término de transferencia de masa en la ecuación (5.1) puede expresarse

$$K_a q_A (S_{O,sat} - x_1) = K_a u (S_{O,sat} - x_1)$$

La energía requerida para aireación puede disminuirse considerablemente utilizando un control de oxígeno disuelto (control DO). Actualmente, en casi la totalidad de las plantas europeas dicho control es standard. Se ha implementado con resultado satisfactorio la utilización de controladores adaptivos o auto-sintonizados, que no dan una estimación explícita de los parámetros $K_L a$ y velocidad de consumo de O_2 , sino que calculan directamente los parámetros del controlador. Se observa que la mayoría de las veces las constantes del controlador adaptivo convergen a un controlador proporcional + integral (PI) [44].

Dado que el objetivo de nuestro trabajo es la DEF, que se realiza a lazo abierto tal como se expresó en la Sección 3.1.1, elegimos un controlador PI cuyas constantes tomamos de la literatura.

Nota 5.4.1 *El algoritmo proporcional + integral + derivativo (PID), utilizado comúnmente en procesos industriales, consiste en la combinación de tres señales de control: proporcional (P), integral (I) y derivativa (D). La elección del modo de combinar dichas acciones depende de cada aplicación en particular, siendo tal vez la más común la acción PI.*

La expresión del algoritmo es la que sigue

$$u = u_0 + K_P e + K_I \int e dt + K_D \frac{de}{dt}$$

donde u_0 es el sesgo de salida del controlador, K_P es la ganancia proporcional, $K_I = 1/T_I$ siendo T_I el tiempo integral, K_D es la ganancia derivativa y e es la diferencia entre el valor de la salida y el valor de referencia o *set point*.

La ganancia proporcional se sintoniza para obtener una estabilidad y amortiguación deseadas en el lazo de control, mientras que el parámetro de ganancia integral es utilizado para remover rápidamente el offset o error de estado estacionario, aunque con el costo de disminuir la estabilidad. Si bien esto puede ser contrarrestado con el agregado de la acción derivativa, ésta presenta el inconveniente de que produce amplificación del ruido.

5.4.7 Cambio de coordenadas. El modelo sin perturbaciones

Para que el modelo tenga su punto de equilibrio en el origen de coordenadas, se refieren las concentraciones a las del estado estacionario, realizando un sencillo cambio de variables

$$z_1 = x_1 - x_1^*, \quad z_2 = x_2 - x_2^*, \quad z_3 = x_3 - x_3^* \quad (5.7)$$

donde x_1^* , x_2^* , x_3^* , son las concentraciones de oxígeno, sustrato y biomasa en estado estacionario, respectivamente.

Finalmente, se obtienen las siguientes ecuaciones

$$\begin{aligned} \dot{z} &= A z + \begin{bmatrix} e \\ e - 1 \\ 1 \end{bmatrix} \mu(z_1, z_2, z_3) + \begin{bmatrix} K_a(S_{O,sat} - z_1 - x_1^*) \\ 0 \\ 0 \end{bmatrix} u \\ y &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} z \end{aligned} \quad (5.8)$$

donde

$$A = \begin{bmatrix} -a & 0 & 0 \\ 0 & -d & c \\ 0 & 0 & -b \end{bmatrix}$$

siendo

$$\begin{aligned} a &= \frac{q_F + q_R}{V} + K_a q_A^* & b &= \frac{q_w}{V} \left(\frac{q_F + q_R}{q_W + q_R} \right) + b_H \\ c &= b_H(1 - f_p) & d &= q_F/V \\ e &= (Y_H - 1)/Y_H \end{aligned}$$

$$\begin{aligned} \mu(z_1, z_2, z_3) = & \\ \mu_H & \left[\left(\frac{z_2 + x_2^*}{K_S + z_2 + x_2^*} \right) \left(\frac{z_1 + x_1^*}{K_{OH} + z_1 + x_1^*} \right) (z_3 + x_3^*) - \right. \\ & \left. - \frac{x_1^* x_2^* x_3^*}{(K_S + x_2^*)(K_{OH} + x_1^*)} \right] \end{aligned} \quad (5.9)$$

Debido a la forma de μ dada por la ecuación (5.9), la parte no lineal del sistema cuenta aún con términos lineales absorbidos en la misma, en consecuencia para incluir estos últimos en la matriz A debe realizarse el desarrollo en serie de Taylor de μ alrededor de $z = 0$. De este modo, los términos lineales de dicho desarrollo pueden incluirse en la parte lineal del sistema.

Entonces se obtiene

$$\begin{aligned} \dot{z} = A z + & \begin{bmatrix} e \\ e - 1 \\ 1 \end{bmatrix} (\mu_1 z_1 + \mu_2 z_2 + \mu_3 z_3) + \\ & + \begin{bmatrix} e \\ e - 1 \\ 1 \end{bmatrix} \mu_R(z_1, z_2, z_3) + \begin{bmatrix} K_a(S_{O,sat} - z_1 - x_1^*) \\ 0 \\ 0 \end{bmatrix} u \end{aligned}$$

Con lo cual el sistema puede ser escrito como

$$\begin{aligned} \dot{z} &= A^* z + \begin{bmatrix} e \\ e - 1 \\ 1 \end{bmatrix} \mu_R(z_1, z_2, z_3) + \begin{bmatrix} K_a(S_{O,sat} - z_1 - x_1^*) \\ 0 \\ 0 \end{bmatrix} u \\ y &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} z \end{aligned} \quad (5.10)$$

donde

$$A^* = \begin{bmatrix} -a + e\mu_1 & e\mu_2 & e\mu_3 \\ (e-1)\mu_1 & -d + (e-1)\mu_2 & c + (e-1)\mu_3 \\ \mu_1 & \mu_2 & -b + \mu_3 \end{bmatrix}$$

Además,

$$\begin{aligned} \mu_1 &= \frac{\mu_H x_2^* x_3^*}{(K_S + x_2^*)(K_{OH} x_1^*)} \cdot \frac{K_{OH}}{(K_{OH} + x_1^*)} & \mu_3 &= \frac{\mu_H x_1^* x_2^*}{(K_S + x_2^*)(K_{OH} x_1^*)} \\ \mu_2 &= \frac{\mu_H x_1^* x_3^*}{(K_S + x_2^*)(K_{OH} x_1^*)} \cdot \frac{K_S}{(K_S + x_2^*)} \end{aligned}$$

y μ_R es el resto, que involucra los términos de segundo orden del desarrollo en serie de Taylor, cuya expresión queda

$$\mu_R = k_1 z_1^2 + k_2 z_2^2 + k_3 z_1 z_2 + k_4 z_1 z_3 + k_5 z_2 z_3 \quad (5.11)$$

donde

$$\begin{aligned} k_1 &= \frac{\mu_H K_{OH} x_2^* x_3^*}{(K_S + x_2^*)(K_{OH} + x_1^*)^3} & k_2 &= \frac{\mu_H K_S x_1^* x_3^*}{(K_{OH} + x_1^*)(K_S + x_2^*)^3} \\ k_3 &= \frac{\mu_H K_S K_{OH} x_3^*}{(K_{OH} + x_1^*)^2 (K_S + x_2^*)^2} & k_4 &= \frac{\mu_H K_{OH} x_2^*}{(K_{OH} + x_1^*)^2 (K_S + x_2^*)} \\ k_5 &= \frac{\mu_H K_S x_1^*}{(K_{OH} + x_1^*)(K_S + x_2^*)^2} \end{aligned}$$

Nótese que la linealización es observable, ya que la matriz de observabilidad

$$\mathcal{O} = \begin{bmatrix} C \\ CA^* \\ C(A^*)^2 \end{bmatrix}$$

tiene rango 3.

5.5 El modelo con perturbaciones

Las ecuaciones anteriores podrían sintetizarse expresando

$$\dot{z} = F(z) + G(z)u = L(z, u)$$

En nuestro caso, las perturbaciones externas pueden asociarse al caudal de alimentación y la concentración de sustrato en la alimentación. Debido a las características de la planta, ambos parámetros pueden presentar un importante rango de variación en función de las fluctuaciones en los sistemas que aportan el agua residual a la misma. Entonces, una descripción más realista de la dinámica del sistema debería incluirlas, siendo

$$\dot{z} = L(z, u, q_F, S_{SF})$$

donde suponemos conocidos y constantes los valores nominales q_{F0} y S_{SF0} .

Entonces, las perturbaciones pueden aproximarse expandiendo $L(z, u, q_F, S_{SF})$ en serie de Taylor alrededor de dichos parámetros nominales [23]. El desarrollo puede concluirse luego del término lineal, dejando de lado perturbaciones de segundo orden y superior, y se obtiene

$$\begin{aligned} L(z, u, q_F, S_{SF}) &= L(z, u, q_{F0}, S_{SF0}) + \\ &+ \frac{\partial L(z, u, q_F, S_{SF})}{\partial q_F} \Delta q + \frac{\partial L(z, u, q_F, S_{SF})}{\partial S_{SF}} \Delta S \end{aligned}$$

donde Δq y ΔS denotan las desviaciones del caudal y la concentración de sustrato de sus valores nominales.

En consecuencia, la matriz de distribución de las perturbaciones $E(z)$ queda dada por

$$\begin{aligned} E(z) &= \begin{bmatrix} \frac{\partial L(z, u, q_F, S_{SF})}{\partial q_F} & \frac{\partial L(z, u, q_F, S_{SF})}{\partial S_{SF}} \end{bmatrix} \\ &= \begin{bmatrix} \frac{S_{OF} - z_1 - x_1^*}{V} & 0 \\ \frac{S_{SFO} - z_2 - x_2^*}{V} & q_{F0} \\ \frac{X_{HF} - f(z_3 - x_3^*)}{V} & 0 \end{bmatrix} \end{aligned} \quad (5.12)$$

Finalmente, se obtiene el siguiente sistema, que representa al modelo con perturbaciones

$$\begin{aligned} \dot{z} &= A^* z + \begin{bmatrix} e \\ e - 1 \\ 1 \end{bmatrix} \mu_R(z_1, z_2, z_3) + \\ &+ \frac{1}{V} \begin{bmatrix} S_{OF} - z_1 - x_1^* & 0 \\ S_{SFO} - z_2 - x_2^* & q_{F0} \\ X_{HF} - f(z_3 + x_3^*) & 0 \end{bmatrix} \begin{bmatrix} \Delta Q \\ \Delta S \end{bmatrix} + \begin{bmatrix} K_a(S_{O,sat} - z_1 - x_1^*) \\ 0 \\ 0 \end{bmatrix} u \\ y &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} z \end{aligned} \quad (5.13)$$

donde A^* es la matriz de la ecuación (5.10), con

$$\begin{aligned} a &= \frac{q_{F0} + q_R}{V} + K_a q_A^* & b &= \frac{q_w}{V} \left(\frac{q_{F0} + q_R}{q_w + q_R} \right) + b_H \\ c &= b_H (1 - f_p) & d &= q_{F0}/V \\ e &= (Y_H - 1)/Y_H & f &= \frac{q_w}{q_w + q_R} \\ q_F &= q_{FO} + \Delta_Q & S_{SF} &= S_{SFO} + \Delta_S \end{aligned}$$

y μ_R fue definida en la Sección anterior.

El modelo propuesto se fundamenta en una versión simplificada del ASM1 (ver Sección 2.3), que es de utilidad a los efectos de la DEF. Para otras aplicaciones de los modelos standard, véanse por ejemplo [34], [57], ya mencionados en dicha sección.

5.6 Algunas fallas típicas en las plantas de tratamiento de efluentes

Una planta de tratamiento de efluentes es un sistema muy complejo, que incluye gran cantidad de equipamiento. El sistema puede ser monitoreado por sensores en línea, instrumentos analíticos, análisis de laboratorio y puede ser observado por un operador. Todas estas mediciones y observaciones constituyen la base para acciones de control u operacionales. En una planta de gran tamaño hay, literalmente, cientos de sensores disponibles. Sólo algunos de ellos miden la calidad del agua, mientras que muchos monitorean variados parámetros físicos.

En la discusión del *diagnóstico* asumiremos que ciertas mediciones se encuentran de alguna manera disponibles. Los sistemas de adquisición de datos pueden recolectar una gran cantidad de ellos, pero existen relativamente pocos eventos relevantes a tener en cuenta. Por lo tanto, los datos de todas las mediciones deben ser transformados en descripciones significativas de lo que está sucediendo en la planta. El problema de la medición no es trivial, ya que por largo tiempo la falta de instrumentos ha sido un obstáculo para la operación de estas plantas. Sin embargo, actualmente se observan importantes desarrollos en la instrumentación. Otro aspecto a tener en cuenta es que la concentración de sustrato y biomasa sólo pueden determinarse por medio de mediciones infrecuentes (análisis de laboratorio de una muestra del proceso), por lo cual estos valores no están disponibles

instantáneamente a los efectos del control o la detección.

Por otra parte, el conocimiento del comportamiento real de las bacterias es aún limitado, si bien se conocen complejos modelos para describir la dinámica del sistema. En consecuencia, la detección precoz de cualquier cambio en el comportamiento del proceso tendrá un valor significativo en la operación.

Con respecto a la detección y el diagnóstico, durante la operación *normal* de la planta las relaciones “causa-efecto” se encuentran bien definidas. Aun cuando una variable presente grandes desviaciones, el sistema de control ejercerá la acción correctiva correspondiente. Por el contrario, en una situación *anormal* hay muchas más opciones, ya que no es tan evidente cuáles son las variables a manipular o los lazos de control a utilizar. Por ejemplo, no debe accionarse de la misma manera si se produce una falla en el equipamiento que si el proceso se ve afectado por un problema en la floculación del lodo.

La interpretación y análisis de los datos es un aspecto importante en la operación de cualquier planta. Aquí definiremos como *detección* a la combinación de observaciones del proceso y mediciones, análisis de datos e interpretación para detectar comportamientos anormales o *efectos* y el aislamiento de fallas. En este contexto hay varios temas de interés, incluyendo descripciones de estados, tendencias, cambios del proceso y descripción de fallas. La detección no comprende explicaciones ni causas de los comportamientos. Por su parte el *diagnóstico* involucra el análisis de los efectos para identificar y clasificar causas probables. Esto generalmente comprende cálculos iterativos y testeo de hipótesis basados en relaciones conocidas causa-efecto. El *informe* consiste en la síntesis de las estrategias que pueden ser implementadas por los operadores de planta para eliminar las “causas” y retornar el proceso a sus condiciones operativas normales. El objetivo final de la detección y diagnóstico es hallar sistemas de alarma temprana para cambios en el proceso. Esto es particularmente importante en un sistema biológico, donde algunos de los cambios pueden no ser muy evidentes y aumentar gradualmente hasta transformarse en un serio problema operacional.

Enumeraremos a continuación algunas de las fallas que podrían aparecer en nuestra planta:

- Medición errónea de la concentración de oxígeno disuelto por ruptura de la membrana del sensor correspondiente.

- Determinación incorrecta de las concentraciones de sustrato y biomasa en las mediciones infrecuentes (no se encuentran modelizadas en este trabajo).
- Problemas de sedimentación, como consecuencia de los cuales se observa una elevada concentración de sustrato en la corriente de salida del clarificador.
- Disminución de la concentración de biomasa, debido a la aparición de algún material tóxico que destruya o inhiba la reproducción de los microorganismos.
- Bloqueo de una válvula. Por ejemplo, en la corriente de entrada, lo cual produciría una disminución del caudal de alimentación, o en la corriente de reflujo, como consecuencia de lo cual disminuiría la biomasa activa en el biorreactor.

La importancia de la determinación de las verdaderas causas de una falla se pone de manifiesto en el siguiente ejemplo. Sabemos que en un biorreactor con control de oxígeno el caudal de aire es proporcional a la carga biodegradable en el sistema. Supongamos que el caudal de aire demandado para mantener la concentración de O_2 en el valor deseado disminuye abruptamente. El sistema de control sólo puede detectar que hay un decaimiento veloz en la carga a la planta. Sin embargo, los motivos pueden ser: la aparición de algún material tóxico que destruya o inhiba la reproducción de los microorganismos (falla en el proceso), o una disminución real en la concentración de sustrato de la alimentación (perturbación externa). Las acciones a desarrollar para retornar a la operación “normal” de la planta son esencialmente diferentes en ambos casos. Esto pone en evidencia la importancia de la robustez frente a perturbaciones en el sistema de DEF.

Capítulo 6

Detección de fallas en una planta de tratamiento de efluentes

En este capítulo, aplicaremos la teoría de la DEF y los ONL a un sistema de tratamiento de efluentes cuyas características se describieron en el capítulo anterior. En primer lugar, desarrollaremos el diseño de tres observadores propuestos para el modelo, comparando su desempeño en función de los requerimientos de la DEF. Presentaremos además las simulaciones que ilustran el funcionamiento de cada uno de ellos.

En la segunda sección, analizaremos en detalle la robustez de estos observadores frente a perturbaciones. Cuando no se cumplan las condiciones para el rechazo o desacoplamiento total de las mismas, apelaremos a la posibilidad de su atenuación a un nivel predefinido.

Por último, estudiaremos el problema de la DEF en la planta, presentando un posible esquema para la detección de las fallas más comunes en el sistema.

6.1 Observadores no lineales para el modelo en estudio

Estudiaremos a continuación el diseño de observadores no lineales robustos para el modelo propuesto en la Sección 5.4.6. Como se discutió en capítulos

anteriores, la robustez frente a perturbaciones es un requisito esencial para aquellos observadores cuyo objetivo final es la detección y aislamiento de fallas. Veremos además que ciertas características inherentes a los sistemas no lineales dificultan la generalización de algunos resultados, así como restringen la aplicabilidad de técnicas clásicas muy conocidas que, siendo efectivas para una clase de SNL, no lo son para otros.

En nuestro análisis tendremos en cuenta que en un SNL la observabilidad es, en general, una propiedad local, dependiente de un punto de operación. Además, lograr la convergencia exponencial del error de estimación requiere el cumplimiento de distintas hipótesis, en ocasiones muy restrictivas. Si a esto agregamos el requerimiento de que el residuo posea sensibilidad máxima frente a las fallas y mínima (en el caso ideal, nula) frente a las perturbaciones, tendremos un panorama de la complejidad del problema en estudio.

Propondremos en esta sección tres observadores para la planta de tratamiento de efluentes, y compararemos luego su desempeño en función de los requerimientos antes planteados. Estos observadores son:

- *Observador de orden reducido*
- *Observador de estructura variable*
- *Observador del tipo filtro de Kalman extendido (TFKE)*

6.1.1 Observador de orden reducido

Desarrollaremos aquí el diseño de un observador de orden reducido, de tipo identidad (ver la Sección 3.4.1). Se construye como una copia de la dinámica del modelo, adicionándole un término de corrección que es una función no lineal de las mediciones. La técnica utilizada puede considerarse como una extensión del método para sistemas no lineales con salida lineal [16], [53].

Para la planta de tratamiento de efluentes, se puede diseñar un observador de orden reducido empleando la concentración de oxígeno que se mide directamente, con lo cual sólo se hace necesario estimar las concentraciones de sustrato y biomasa. Este tipo de observador permite reducir la cantidad de operaciones necesarias para su implementación, y en muchos casos se

simplifica el análisis de estabilidad.

Para el sistema (5.13), se propone el observador

$$\begin{aligned}\dot{\hat{z}}_2 &= (e-1)\mu_1 z_1 + [-d + (e-1)\mu_2] \hat{z}_2 + [c + (e-1)\mu_3] \hat{z}_3 - \\ &\quad - (1-e)\mu_R(z_1, \hat{z}_2, \hat{z}_3) + g_1 \epsilon \\ \dot{\hat{z}}_3 &= \mu_1 z_1 + \mu_2 \hat{z}_2 + (-b + \mu_3) \hat{z}_3 + \mu_R(z_1, \hat{z}_2, \hat{z}_3) + g_2 \epsilon\end{aligned}\quad (6.1)$$

donde g_1 y g_2 son parámetros constantes que deben ser diseñados para garantizar la convergencia a cero del error de estimación, y

$$\epsilon = \dot{z}_1 - \dot{\hat{z}}_1$$

Notemos que la dinámica del observador es copia de la dinámica del sistema más un término de corrección proporcional a ϵ . La expresión elegida para ϵ permite la reducción del orden; pero un término de corrección que utiliza derivadas de las mediciones presenta el inconveniente de que las mismas no se obtienen por medición directa, ni tampoco es conveniente calcularlas a partir de derivadas numéricas, ya que ello originaría estimaciones ruidosas. Para evitar esto, realizaremos un cambio de variables en el observador:

$$\begin{aligned}\hat{z}_2 &= w_2 + g_1 z_1 \\ \hat{z}_3 &= w_3 + g_2 z_1\end{aligned}$$

Entonces, las ecuaciones quedan

$$\begin{aligned}\dot{\hat{z}}_1 &= (-a + e\mu_1) z_1 + e\mu_2 \hat{z}_2 + e\mu_3 \hat{z}_3 + \\ &\quad + e\mu_R(z_1, \hat{z}_2, \hat{z}_3) + K_a(S_{O,sat} - z_1 - x_1^*)u \\ \dot{w}_2 &= (e-1)\mu_1 z_1 + [-d + (e-1)\mu_2] \hat{z}_2 + [c + (e-1)\mu_3] \hat{z}_3 - \\ &\quad - (1-e)\mu_R(z_1, \hat{z}_2, \hat{z}_3) + g_1 \dot{\hat{z}}_1 \\ \dot{w}_3 &= \mu_1 z_1 + \mu_2 \hat{z}_2 + (-b + \mu_3) \hat{z}_3 + \mu_R(z_1, \hat{z}_2, \hat{z}_3) + g_2 \dot{\hat{z}}_1\end{aligned}\quad (6.2)$$

Observemos que en las ecuaciones del observador y como consecuencia del cambio de variables, en lugar del término de corrección ϵ aparece $\dot{\hat{z}}_1$, que se obtiene sin derivar de la primera ecuación de (6.2).

Se define el error como

$$\tilde{e} = \begin{bmatrix} \tilde{e}_2 \\ \tilde{e}_3 \end{bmatrix} = \begin{bmatrix} z_2 - \hat{z}_2 \\ z_3 - \hat{z}_3 \end{bmatrix}\quad (6.3)$$

Su dinámica queda dada por

$$\begin{aligned} \dot{\tilde{e}} &= A\tilde{e} + \begin{bmatrix} e-1 \\ 1 \end{bmatrix} [\mu_R(z_1, z_2, z_3) - \\ &- \mu_R(z_1, \hat{z}_2, \hat{z}_3)] - \begin{bmatrix} g_1 \\ g_2 \end{bmatrix} \epsilon \end{aligned} \quad (6.4)$$

donde

$$A = \begin{bmatrix} -d + (e-1)\mu_2 & c + (e-1)\mu_3 \\ \mu_2 & -b + \mu_3 \end{bmatrix}$$

Reemplazando ϵ en (6.4), se obtiene

$$\dot{\tilde{e}} = A\tilde{e} + G[\mu_R(z_1, z_2, z_3) - \mu_R(z_1, \hat{z}_2, \hat{z}_3)] \quad (6.5)$$

con $G = [-1 + e - eg_1, 1 - eg_2]^T$

La solución de la ecuación anterior es

$$\tilde{e}(t) = e^{At}\tilde{e}_0 + \quad (6.6)$$

$$\int_0^t e^{A(t-\tau)} G[\mu_R(z_1, z_2, z_3) - \mu_R(z_1, \hat{z}_2, \hat{z}_3)] d\tau \quad (6.7)$$

La matriz A tiene autovalores λ_1 y λ_2 reales, negativos y distintos, y los autovectores correspondientes son $[\lambda_1 \ 1]$ y $[\lambda_2 \ 1]$. Entonces, llamando $V(\lambda)$ a la matriz (de Vandermonde)

$$V(\lambda) = \begin{bmatrix} \lambda_1 & 1 \\ \lambda_2 & 1 \end{bmatrix}$$

se puede diagonalizar la matriz e^{At} como sigue

$$e^{At} = V^{-1}(\lambda) e^{\Lambda t} V(\lambda)$$

donde $\Lambda = \text{diag}(\lambda_1 \ \lambda_2)$ y $e^{\Lambda t} = \text{diag}(e^{\lambda_1 t} \ e^{\lambda_2 t})$.

Sustituyendo estas expresiones en (6.6) se obtiene

$$\begin{aligned} \tilde{e}(t) &= V^{-1}(\lambda) e^{\Lambda t} V(\lambda) \tilde{e}_0 + \\ &\int_0^t V^{-1}(\lambda) e^{\Lambda(t-\tau)} V(\lambda) e^{A(t-\tau)} G[\mu_R(z_1, z_2, z_3) - \mu_R(z_1, \hat{z}_2, \hat{z}_3)] d\tau \end{aligned}$$

Analizaremos la norma del error, para lo cual tengamos en cuenta que

- $\| e^{\Lambda t} \| = e^{\lambda_1 t}$,
pues $\lambda_1 > \lambda_2$.
- $\| \mu_R(z, t) - \mu_R(\hat{z}, t) \| \leq L \| z - \hat{z} \|$
ya que se ha considerado que μ_R satisface la condición de ser globalmente Lipschitz en la variable z con constante L .

Utilizando además el lema de Gronwall [16], [37] (Ver Apéndice B), se llega a

$$\| \tilde{e}(t) \| \leq \| V^{-1}(\lambda) \| \| V(\lambda) \| \| \tilde{e}_0 \| e^{(\lambda_1 + \| V^{-1}(\lambda) \| \| V(\lambda)G \| L)t} \quad (6.8)$$

Por tanto, para que el observador converja debe elegirse G de modo que se verifique

$$-\alpha = \lambda_1 + \| V^{-1}(\lambda) \| \| V(\lambda)G \| L < 0, \quad (6.9)$$

con lo cual

$$\| \tilde{e}(t) \| \leq k_1 e^{-\alpha t} \| \tilde{e}_0 \| \quad (6.10)$$

donde

$$\alpha > 0$$

$$k_1 = \| V^{-1}(\lambda) \| \| V(\lambda) \|$$

y la convergencia del error es exponencial.

En síntesis, α corresponde a un valor de G con el cual se verifica la ecuación (6.9), y teniendo en cuenta las expresiones anteriores,

$$\begin{aligned} \| V(\lambda)G \| = & \left\{ [1 - \lambda_1 + e(\lambda_1 - \lambda_1 g_1 + g_2)]^2 + \right. \\ & \left. + [1 - \lambda_2 + e(\lambda_2 - \lambda_2 g_1 - g_2)]^2 \right\}^{1/2} \end{aligned} \quad (6.11)$$

Reemplazando en (6.9) y elevando al cuadrado, se obtiene

$$\begin{aligned} & [1 - \lambda_1 + e(\lambda_1 - \lambda_1 g_1 + g_2)]^2 + [1 - \lambda_2 + \\ & + e(\lambda_2 - \lambda_2 g_1 - g_2)]^2 < \left(\frac{\lambda_1}{\| V^{-1}(\lambda) \| L} \right)^2 \end{aligned} \quad (6.12)$$

El objetivo de diseño se reduce, entonces, a la elección de los valores adecuados de g_1 y g_2 para que se verifique la desigualdad anterior.

Tabla 6.1: Valores para la planta típica

SIMBOLO	VALOR
q_F	50 [Ml/d]
S_{SF}	150 [mg/l]
S_{OF}	1 [mg/l]
X_{HF}	5 [mg/l]
q_A	9.072 [Ml/d]
q_R	50 [Ml/d]
q_W	1.615 [Ml/d]
V	12.5 [Ml]
μ_H	3.733 [1/día]
K_S	20 [mg/l]
K_{OH}	0.2 [mg/l]
Y_H	0.6
b_H	0.4 [1/día]
f_p	0.1
$S_{O,sat}$	10 [mg/l]
K_a	5.955 [1/Ml]

6.1.1.1 Resultados de aplicación

A continuación analizaremos el resultado obtenido de la aplicación del procedimiento de diseño desarrollado anteriormente a un proceso de tratamiento de efluentes. Los parámetros para el modelo fueron tomados de [44] para una planta típica, y se presentan en la Tabla 6.1. Estos datos serán utilizados en todas las simulaciones que presentemos a partir de aquí.

Realizando entonces los cálculos correspondientes, de la expresión (6.12) se obtiene

$$(1.7095 - 0.2838 g_1 + g_2)^2 + (286.7042 - 114.2817 g_1 + \frac{2}{3} g_2)^2 < 5.59 \cdot 10^{-4} \quad (6.13)$$

Esta se verifica con:

$$g_1 = 2.5029 \quad y \quad g_2 = -0.9991$$

Analizando en forma más exhaustiva este resultado, se observa que estos son “casi” los únicos valores de g_1 y g_2 que satisfacen la desigualdad (6.13),

es decir, el rango de variación de ambas ganancias es muy pequeño.

En efecto, si tomamos $g_1 = 2.502$ y $g_2 = -0.999$ no se verifica dicha desigualdad, es decir que el observador no tolera una variación de las ganancias del orden de 10^{-4} . Se observa que, para los valores dados, $\|V(\lambda)G\|^2 < 10^{-4}$ (aproximadamente), y en consecuencia los elementos de $V(\lambda)G$ deben ser del orden de 10^{-2} . Ahora bien, recordemos que

$$G = [e - 1 - e g_1, 1 - e g_2]^T \quad (6.14)$$

en consecuencia ni bien e varíe en alrededor de 10^{-2} podría no satisfacerse la ecuación (6.13).

Físicamente, el parámetro e es una función del coeficiente de rendimiento, con lo cual es esperable que pueda presentar alguna variación respecto de su valor nominal. El análisis previo conduce a que, ni bien e se aparte de dicho valor nominal el observador se torna inestable, ya que deja de verificarse la desigualdad (6.13). De ahí la inutilidad de este observador que, aunque en apariencia funcione correctamente, y de hecho lo hace en las simulaciones para valores nominales, posee robustez prácticamente nula. Esto refuerza la idea, planteada antes, de que en el diseño de observadores para sistemas no lineales deben estudiarse cuidadosamente las características particulares de cada proceso, las que en muchos casos limitan la aplicabilidad de técnicas usuales para cierta clase de sistemas no lineales.

6.1.2 Observador de estructura variable

Se propone a continuación un observador de orden completo y de estructura variable. Su principal ventaja es que no requiere del conocimiento exacto de las no linealidades del sistema, lo que le confiere una robustez inherente [60]. Esto se logra utilizando técnicas que se desarrollan en la teoría de los sistemas de estructura variable.

En cuanto a sus desventajas, veremos que el modelo debe satisfacer ciertas hipótesis adicionales, lo que condiciona su aplicabilidad.

Para el proceso en estudio, el observador de estructura variable (OEV) queda dado por

$$\dot{\hat{z}} = A^* \hat{z} + S(\hat{z}, y, \rho) + \bar{B}u \quad (6.15)$$

con el error definido como $\tilde{e} = z - \hat{z}$.

Además

$$\bar{B} = [K_a(S_{O,sat} - z_1 - x_1^*), 0, 0]^T$$

$$S(\hat{z}, y, \rho) = \begin{cases} \frac{P^{-1}C^T C \tilde{e} \rho}{\|C \tilde{e} \rho\|} \rho(t, u) & \text{si } \|C \tilde{e} \rho\| > \delta \\ \frac{P^{-1}C^T C \tilde{e} \rho}{\delta} \rho(t, u) & \text{si } \|C \tilde{e} \rho\| \leq \delta \end{cases} \quad (6.16)$$

Observemos que se verifican las siguientes hipótesis, que posibilitan que el sistema (6.15) sea, en efecto, un buen observador para nuestro sistema:

H1) El par (A^*, C) es observable.

H2) Existe una función escalar positiva $\rho(t, u)$ que es una cota uniforme del término no lineal de la planta. Debido a la forma especial $B h(t, z) = B \mu_R(z_1, z_2, z_3)$ de dicho término, con $B = [e, e - 1, 1]^T$, se cumple que:

$$\|h(t, z)\| = \|\mu_R(z_1, z_2, z_3)\| \leq \rho(t, u), \quad \forall t.$$

H3) La matriz P (simétrica y positiva definida) debe satisfacer ciertas condiciones de forma ("matching conditions"):

$$\begin{aligned} A^{*T}P + PA^* &= -\bar{Q} \\ PB &= C^T \end{aligned} \quad (6.17)$$

para alguna \bar{Q} simétrica.

Estas condiciones suelen ser *a-priori* restrictivas, aunque el lema de Meyer-Kalman -Yacubovich [41] (Ver **Apéndice A**), es de gran ayuda para saber si se pueden cumplir y además es instrumental para hallar \bar{Q} , P .

El lema establece que la condición de que la transferencia definida por (A^*, B, C) sea estrictamente real positiva (o sea, $\Re\{C(j\omega I - A^*)^{-1}B\} > 0$) es necesaria y suficiente para que exista Q simétrica, definida positiva, de modo que se verifiquen simultáneamente las condiciones de forma (6.17), con

$$\bar{Q} = \gamma Q + qq^T$$

donde γ y q son una constante y un vector que pueden calcularse a partir de (A^*, B, C) .

La dinámica del error para el observador de (6.15) queda dada por

$$\dot{\tilde{e}} = A^* \tilde{e} - S(\hat{z}, y, \rho) - P^{-1} C^T \mu_R(\hat{z}_1, \hat{z}_2, \hat{z}_3) \quad (6.18)$$

Consideremos la siguiente función de Lyapunov, definida positiva

$$V(\tilde{e}) = \tilde{e}^T P \tilde{e}$$

donde P fue definida en H3).

La derivada de dicha función con respecto al tiempo está dada por

$$\dot{V}(\tilde{e}) = -\tilde{e}^T \bar{Q} \tilde{e} - 2 \|C \tilde{e}\| \rho + 2 \tilde{e}^T C^T h(t, z) \quad (6.19)$$

Tomando la norma euclídea en el último término de (6.19) y teniendo en cuenta H2), se obtiene

$$\dot{V}(\tilde{e}) \leq -\tilde{e}^T \bar{Q} \tilde{e} - 2 \|C \tilde{e}\| \rho + 2 \|C \tilde{e}\| \rho < 0 \quad (6.20)$$

Entonces, $\lim_{t \rightarrow \infty} \tilde{e}(t) = 0$, con lo cual la diferencia entre el estado y su estimación tiende a cero asintóticamente. Para conocer la velocidad de convergencia del observador, consideremos

$$\frac{-\dot{V}(\tilde{e})}{V(\tilde{e})} \geq \frac{\tilde{e}^T \bar{Q} \tilde{e}}{\tilde{e}^T P \tilde{e}}$$

es decir que

$$V(\tilde{e}(t)) \leq V(\tilde{e}_0, t_0) \tilde{e}^{-\eta(t-t_0)}$$

siendo η el mínimo autovalor de $P^{-1} \bar{Q}$ [60].

Nótese que de los dos observadores presentados en [60] hemos utilizado el de capa límite. Este permite evitar la discontinuidad en el término $S(\hat{z}, y, \rho)$, así como también el fenómeno conocido como “chattering”, común en los sistemas de estructura variable. El análisis de convergencia del observador es análogo y se concluye que, eligiendo en (6.16) un valor de δ suficientemente pequeño, el mismo converge exponencialmente con velocidad de decaimiento proporcional a $e^{-\eta t}$, donde η es el mínimo autovalor de $P^{-1} \bar{Q}$.

6.1.2.1 Simulaciones y resultados para el OEV

Reemplazando valores (de Tabla 6.1) en (6.15) y (6.16), se obtiene, con $\delta/\rho = 6 \cdot 10^{-6}$

$$\begin{aligned} \dot{\hat{z}} = & \begin{bmatrix} -69.2304 & -66.969 & -0.4179 \\ -18.0165 & -171.4225 & -0.6847 \\ 10.8099 & 100.4535 & -0.0235 \end{bmatrix} \hat{z} + S(\hat{z}, y, \rho) \\ & + \begin{bmatrix} 5.955(6.9573 - z_1) \\ 0 \\ 0 \end{bmatrix} u \end{aligned}$$

$$S(\hat{z}, y, \rho) = \begin{cases} 150 \begin{bmatrix} 2/3 \\ 5/3 \\ -1 \end{bmatrix} \frac{z_1 - \hat{z}_1}{\|z_1 - \hat{z}_1\|} & \text{si } \|z_1 - \hat{z}_1\| > \frac{\delta}{\rho} \\ 150 \begin{bmatrix} 2/3 \\ 5/3 \\ -1 \end{bmatrix} \frac{z_1 - \hat{z}_1}{\delta/150} & \text{si } \|z_1 - \hat{z}_1\| \leq \frac{\delta}{\rho} \end{cases}$$

De este modo, el observador verifica las hipótesis:

- H1) El par (A^*, C) es observable.
- H2) El término no lineal de la planta $B\mu_R(z_1, z_2, z_3)$ se encuentra acotado, siendo

$$\|\mu_R(z_1, z_2, z_3)\| \leq 150 [1/\text{hora}] \quad \forall t.$$

- H3) Como se expuso antes, el lema de Meyer Kalman Yacubovich permite hallar

$$\bar{Q} = \begin{bmatrix} 30.0375 & -0.0034 & 0.0092 \\ -0.0034 & 0.0008 & 0 \\ 0.0092 & 0 & -0.0008 \end{bmatrix}$$

Cabe notar que la transferencia del triplete (A^*, B, C) no es positiva real. Sin embargo, cambiando el signo de B , sí lo es. Por la forma especial del término no lineal, esto se puede hacer si se cambia μ_R por

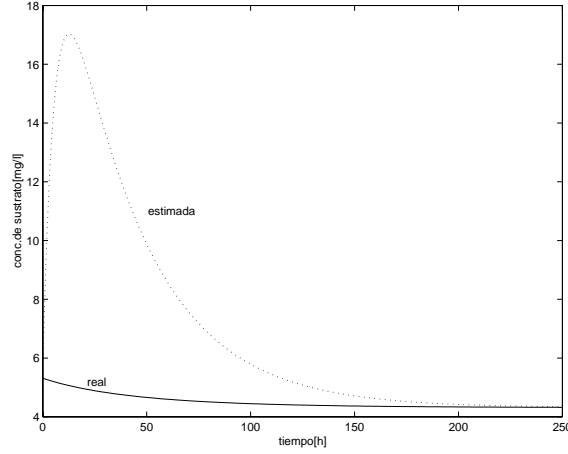


Figura 6.1: Evolución de la concentración de sustrato de la planta (—) y del observador (···).

$-\mu_R$, lo que no afecta la acotación necesaria por $\rho(t, u) = 150[1/hora]$. También se halla

$$P^{-1} = \begin{bmatrix} 1 & 1 & -1 \\ 1 & 10003 & -2 \\ -1 & -2 & 1668 \end{bmatrix}$$

y el error converge eligiendo adecuadamente δ .

Las simulaciones se realizaron con un esquema de control que consta de un controlador PI, usado para seguir una referencia constante en la concentración de oxígeno [44].

En las figuras se observa la performance del observador para las estimación de las concentraciones de sustrato y biomasa. Los valores iniciales del observador son arbitrarios y variaciones de los mismos no producen cambios cualitativos significativos de las respuestas. El error de estimación es aceptable, y la convergencia se produce en el tiempo esperable, acorde con la dinámica del proceso. Nótese que la constante de tiempo de la biomasa es del orden de varios días. No se presenta la variable z_1 , concentración de oxígeno, debido a que el error es muy pequeño y converge rápidamente, por lo cual no es visible en el gráfico.

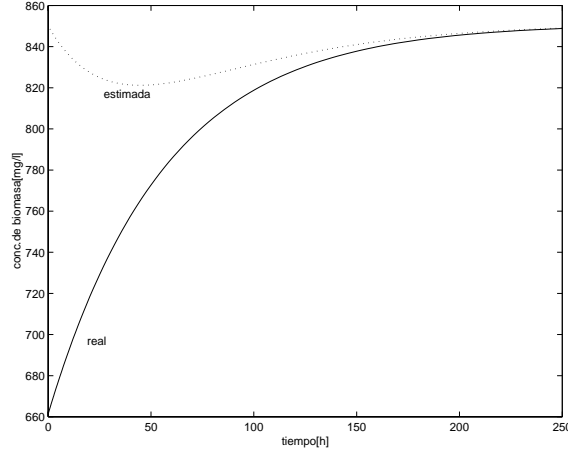


Figura 6.2: Evolución de la concentración de biomasa de la planta (—) y del observador (···).

6.1.3 Observador del tipo Filtro de Kalman Extendido (TF-KE)

Desarrollaremos a continuación un observador del tipo Filtro de Kalman Extendido para el sistema en estudio, siguiendo a [28]. Nuestro objetivo es generar un observador robusto, de modo que las perturbaciones y las incertidumbres en el modelado no afecten al residuo.

El modelo de la planta de tratamiento de aguas residuales puede escribirse como sigue

$$\begin{aligned}\dot{z} &= F(z) + G(z)u + E(z)d \\ y &= h(z)\end{aligned}\tag{6.21}$$

donde

$$\begin{aligned}
 F(z) &= A^*z + \begin{bmatrix} e \\ e-1 \\ 1 \end{bmatrix} \mu_R(z) \\
 G(z) &= \begin{bmatrix} K(S_o - z_1 - x_1^*) \\ 0 \\ 0 \end{bmatrix} \\
 C &= [1 \ 0 \ 0] \\
 E(z) &= \begin{bmatrix} \frac{S_{OF} - z_1 - x_1^*}{V} & 0 \\ \frac{S_{SF0} - z_2 - x_2^*}{V} & q_{F0} \\ \frac{X_{HF} - f(z_3 - x_3^*)}{V} & 0 \end{bmatrix}
 \end{aligned}$$

Para diseñar un observador TFKE tal como en [28], debemos en primer lugar aplicar la transformación que lleva al sistema original a una forma canónica uniformemente observable para toda entrada. Dicha transformación se expresa

$$w = T(z) = \begin{bmatrix} h(z) \\ L_F h(z) \\ L_F^2 h(z) \end{bmatrix}$$

El sistema transformado queda entonces

$$\begin{aligned}
 \dot{w} &= \frac{\partial T}{\partial z} \dot{z} \\
 &= \frac{\partial T}{\partial z} (F(z) + G(z)u + E(z)d) \\
 y &= h(z)
 \end{aligned} \tag{6.22}$$

Si escribimos el jacobiano de la transformación

$$J = \frac{\partial T}{\partial z} = \begin{bmatrix} \frac{\partial h}{\partial z} \\ \frac{\partial L_F h(z)}{\partial z} \\ \frac{\partial L_F^2 h(z)}{\partial z} \end{bmatrix}$$

el sistema puede escribirse

$$\begin{aligned}
\dot{w} &= \begin{bmatrix} L_F h(T^{-1}(w)) \\ L_F^2 h(T^{-1}(w)) \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ L_F^3 h(T^{-1}(w)) \end{bmatrix} + \\
&+ \begin{bmatrix} L_G h(T^{-1}(w)) \\ L_G L_F h(T^{-1}(w)) \\ L_G L_F^2 h(T^{-1}(w)) \end{bmatrix} u + \begin{bmatrix} L_E h(T^{-1}(w)) \\ L_E L_F h(T^{-1}(w)) \\ L_E L_F^2 h(T^{-1}(w)) \end{bmatrix} d \quad (6.23)
\end{aligned}$$

es decir

$$\begin{aligned}
\dot{w} &= Aw + K(w, u) + E(w)d \\
y &= [1 \ 0 \ 0] w \quad (6.24)
\end{aligned}$$

Para el sistema transformado se define un observador como sigue.

$$\begin{aligned}
\dot{\hat{w}} &= A\hat{w} + K(\hat{w}, u) + S_\infty^{-1}(\theta)C^T(y - C\hat{w}) \\
\hat{y} &= [1 \ 0 \ 0] \hat{w} \quad (6.25)
\end{aligned}$$

donde

- $K(\hat{w}, u)$ es uniformemente Lipschitz con constante k .
- $S_\infty(\theta)$ es la solución de la ecuación de Lyapunov

$$A^T S_\infty(\theta) + S_\infty(\theta)A + \theta S_\infty(\theta) - C^T C = 0 \quad (6.26)$$

$$\text{con } A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

El cálculo de $S_\infty(\theta)$ se realiza teniendo en cuenta que (ver demostración en el Apéndice D)

$$(S_\infty(\theta))_{i,j} = (S_\infty(1))_{i,j} \frac{1}{\theta^{i+j-1}} \quad (6.27)$$

con lo cual

$$S_\infty(\theta) = \begin{bmatrix} \frac{1}{\theta} & \frac{-1}{\theta^2} & \frac{1}{\theta^3} \\ \frac{-1}{\theta^2} & \frac{2}{\theta^3} & \frac{-3}{\theta^4} \\ \frac{1}{\theta^3} & \frac{-3}{\theta^4} & \frac{6}{\theta^5} \end{bmatrix}$$

La implementación del observador se realiza en variables originales, debido a su mayor sencillez. Siendo

$$\hat{w} = T(\hat{z}) \Rightarrow \hat{z} = T^{-1}(\hat{w})$$

se cumple que

$$\dot{\hat{z}} = \frac{\partial T^{-1}}{\partial \hat{w}} \dot{\hat{w}} = J^{-1}(\hat{z}) \dot{\hat{w}}$$

Finalmente se obtiene

$$\begin{aligned} \dot{\hat{z}} &= F(\hat{z}) + G(\hat{z})u + J^{-1}(\hat{z}) S_{\infty}^{-1}(\theta) C^T (y - \hat{y}) \\ \hat{y} &= h(\hat{z}) = [1 \ 0 \ 0] \hat{z} = C \hat{z} \end{aligned} \quad (6.28)$$

Nótese que el observador propuesto es muy simple, ya que copia la dinámica del modelo y le agrega un término de corrección que sólo es función de la velocidad de convergencia deseada (θ) y de la dimensión del sistema. De la ecuación (6.28) se ve que tiene la estructura de un filtro de Kalman, cuya ganancia se obtiene de una ecuación de Riccati (6.26). Por esto el nombre de *observador del tipo filtro de Kalman extendido* (TFKE).

6.1.3.1 Análisis de convergencia

Estudiaremos la convergencia del error en el dominio transformado definido previamente, utilizando la siguiente función de Lyapunov

$$V(e) = e^T S_{\infty}(\theta) e$$

Se observa que

$$\dot{V}(e) = 2 e^T S_{\infty}(\theta) \dot{e} \quad (6.29)$$

Siguiendo un razonamiento análogo al de [28], se obtiene

$$\begin{aligned} \dot{V}(e) &= 2 e^T S_{\infty}(\theta) [A - S_{\infty}^{-1}(\theta) C^T C]e + \\ &\quad + 2 e^T S_{\infty}(\theta) [K(w, u) - K(\hat{w}, u)] \\ &\quad + 2 e^T S_{\infty}(\theta) E(w) d \end{aligned} \quad (6.30)$$

y entonces

$$\begin{aligned} \dot{V}(e) &= -\theta e^T S_{\infty}(\theta) e - e^T C^T C e \\ &\quad + 2 e^T S_{\infty}(\theta) [K(w, u) - K(\hat{w}, u)] \\ &\quad + 2 e^T S_{\infty}(\theta) E(w) d \end{aligned} \quad (6.31)$$

Definiendo

$$\|e\|_S^2 = e^T S_\infty(\theta) e$$

y aplicando la desigualdad de Schwartz, se llega a

$$\begin{aligned} \dot{V}(e) &\leq -\theta \|e\|_S^2 \\ &+ 2 \|e\|_S \|K(w, u) - K(\hat{w}, u)\|_S \\ &+ 2 e^T S_\infty(\theta) E(w) d - \|Ce\|^2 \end{aligned} \quad (6.32)$$

Como $K(w, u)$ es uniformemente Lipschitz en la variable w con respecto a u con constante k (nótese que u es acotada, $|u(t)| < u_0$)

$$\begin{aligned} \dot{V}(e) &\leq -\theta \|e\|_S^2 + \frac{2S^{1/2}kn^2}{\theta} \|e\| \|e\|_S \\ &+ 2 e^T S_\infty(\theta) E(w) d - \|e\|_C^2 \end{aligned} \quad (6.33)$$

donde n es el orden del sistema, k es la constante de Lipschitz de K y $S = \sup_{i,j} |S(1)_{i,j}|$.

La convergencia del observador TFKE se produce si $\dot{V}(e) < 0$, ya que esto asegura que $e = 0$ es un punto de equilibrio del modelo dinámico de error. Observemos que la existencia de perturbaciones plantea una dificultad adicional para lograr esta condición.

6.1.3.2 Simulaciones para el observador TFKE

Analizaremos a continuación el resultado obtenido con el observador en la planta de tratamiento de efluentes, ilustrando su comportamiento mediante simulación.

Las simulaciones se realizaron, como antes, con un esquema de control que utiliza un controlador PI para seguir una referencia constante en la concentración de oxígeno [44].

En las figuras 6.3 y 6.4 se observa la performance del observador en la estimación de las concentraciones de sustrato y biomasa. El error de estimación es comparativamente pequeño y la convergencia se produce en un tiempo apreciablemente menor que para otros observadores, como por ejemplo el de estructura variable [55]. No se presenta la variable z_1 , concentración

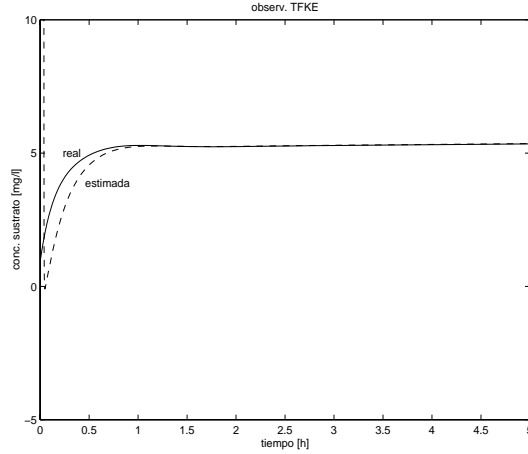


Figura 6.3: Evolución de la concentración de sustrato-referida al estado estacionario- de la planta y del observador (*estimada*).

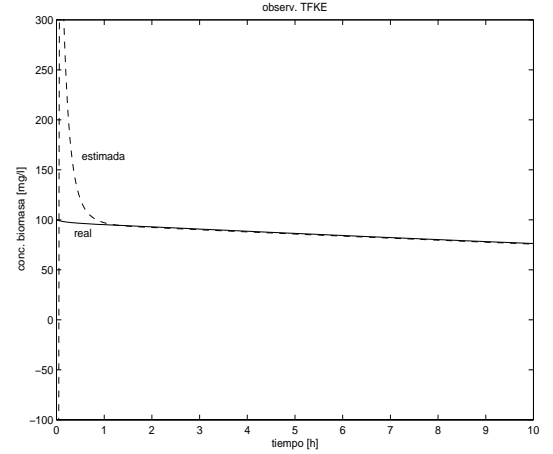


Figura 6.4: Evolución de la concentración de biomasa-referida al estado estacionario- de la planta y del observador (*estimada*).

de oxígeno, debido a que el error es muy pequeño y converge rápidamente, por lo cual no es visible en el gráfico.

El comportamiento de este observador parece notablemente superior, por lo cual lo adoptamos para indagar sobre su desempeño frente a perturbaciones externas.

6.2 Estudio de la robustez del observador TFKE. Atenuación de perturbaciones.

En función de nuestro objetivo, la detección de fallas, sería deseable hallar un observador tal que permita detectar y aislar una falla sin “confundirla” con las perturbaciones esperables en el sistema, como por ejemplo, la variación en el caudal de entrada y en la concentración de sustrato en la alimentación. Las condiciones para lograr que dichas perturbaciones se desacoplen, es decir, que la salida del sistema-en este caso, el observador- sea completamente independiente de ellas [52], son bastante restrictivas y no se cumplen aquí. Por ello se tratará el problema, menos exigente, de atenuar la influencia de las perturbaciones en el error de observación [33].

6.2.1 Modelo de error

Analizaremos ahora el modelo de error de nuestro sistema, en el dominio transformado definido precedentemente. Siendo $\dot{e} = \dot{w} - \dot{\hat{w}}$, se tiene

$$\begin{aligned}\dot{e} &= M(e, w, \hat{w}, u) - S_{\infty}^{-1}(\theta) C^T C e + E(w) d \\ r &= H(e)\end{aligned}\tag{6.34}$$

donde

$$\begin{aligned}M(e, z, \hat{w}, u) &= A e + K(w, u) - K(\hat{w}, u) \\ A &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \\ K(w, u) &= \begin{bmatrix} 0 \\ 0 \\ L_F^3 h(T^{-1}(w)) \end{bmatrix} + \begin{bmatrix} L_G h(T^{-1}(w)) \\ L_G L_F h(T^{-1}(w)) \\ L_G L_F^2 h(T^{-1}(w)) \end{bmatrix} u \\ C &= [1 \ 0 \ 0] \\ E(w) &= \begin{bmatrix} L_E h(T^{-1}(w)) \\ L_E L_F h(T^{-1}(w)) \\ L_E L_F^2 h(T^{-1}(w)) \end{bmatrix} \\ d &= \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}\end{aligned}$$

y también definimos $\Gamma = M - S_{\infty}^{-1}(\theta) C^T C e$.

Nótese que para hallar $E(w)$ se aplicó la transformación $T(z)$ a la matriz de entrada de las perturbaciones (Sección 5.5), que aquí llamamos E , obteniéndose una matriz de dimensiones compatibles con d .

En este caso, hemos tomado al residuo r como salida del modelo de error. La variación en el caudal de alimentación de la planta es representada por d_1 , y la variación en la concentración de sustrato por d_2 . De esta manera, queremos atenuar la ganancia entre d y r .

6.2.2 Convergencia del observador TFKE con atenuación de perturbaciones

Como ya se dijo, se desea analizar si es posible reducir el efecto de las entradas desconocidas en el error de observación. En primer lugar, es necesario

establecer un criterio preciso para poder medir la “influencia” de una entrada dada (en este caso, las perturbaciones) en la salida del sistema. Como se vio en el Capítulo 4, en el caso de los sistemas no lineales, una noción asociada a dicho concepto es la llamada “*ganancia* \mathcal{L}_2 ”. El sistema descrito según (6.34) tiene ganancia \mathcal{L}_2 menor o igual que γ si para toda $d(\cdot) \in \mathcal{L}_2$ la respuesta $e(s, 0, d(\cdot))$ existe y satisface $\forall t > 0$

$$\int_0^t \| H(e(s, 0, d(\cdot))) \|^2 ds \leq \gamma^2 \int_0^t \| d(s) \|^2 ds$$

donde $H(e)$ es el residuo, es decir, la salida del modelo de error.

Tal como se vio en el Capítulo 4, una condición suficiente para que se cumpla lo anterior está dada por la existencia de una función, en nuestro caso $V(e)$, propia y definida positiva, que satisfaga una desigualdad de Hamilton-Jacobi. En dicho capítulo se observa que el sistema tendrá un equilibrio asintóticamente estable en $e = 0$, con ganancia \mathcal{L}_2 menor o igual que γ , si se verifica

$$L_\Gamma V(e) + \frac{1}{4\gamma^2} [L_E V(e)]^2 + [H(e)]^2 < 0 \quad (6.35)$$

donde:

$$L_\Gamma V(e) = 2 e^T S_\infty(\theta) \Gamma(e) \quad (6.36)$$

$$L_E V(e) = 2e^T S_\infty(\theta) E(w) \quad (6.37)$$

$$[H(e)]^2 = e^T e = \| e \|^2 \quad (6.38)$$

La condición (6.35) es equivalente a [33]

$$\frac{\partial V}{\partial e} [\Gamma(e) + E(w) d] + \| e \|^2 - \gamma^2 \| d \|^2 < 0 \quad \forall d \quad (6.39)$$

Para expresar la desigualdad (6.35) en función de la norma del error, deben tenerse en cuenta (6.36), (6.37) y (6.38), llegándose a

$$\begin{aligned} & -\theta \| e \|_S^2 + \frac{2S^{1/2}kn^2}{\theta} \| e \| \| e \|_S + \| e \|^2 + \\ & + \frac{1}{\gamma^2} e^T S_\infty(\theta) E(w) E^T(w) S_\infty(\theta) e < 0 \end{aligned}$$

Consideremos además que

$$\| e \|^2 \leq \frac{1}{\lambda_m} \| e \|_S^2$$

siendo λ_m el mínimo autovalor de $S_\infty(\theta)$. Observemos que como (C, A^*) es un par observable, existe $\delta > 0$ tal que $S_\infty(\theta) \geq \delta I$ y entonces $\lambda_m\{S_\infty(\theta)\} \geq \delta > 0$. En consecuencia, se llega a la condición

$$\begin{aligned} & \| e \|_S^2 \left(-\theta + \frac{2(S\delta)^{1/2}kn^2}{\theta} + \delta \right) + \\ & + \frac{1}{\gamma^2} e^T S_\infty(\theta) E(w) E^T(w) S_\infty(\theta) e < 0 \end{aligned} \quad (6.40)$$

El próximo paso consiste en la obtención de una cota para $E(w)$ de modo que pueda realizarse un nuevo diseño del observador, posibilitando la atenuación de perturbaciones a un nivel predefinido.

La expresión anterior debe escribirse en función de la norma de $E(w)$, con el objetivo de recalculer el valor de θ para el observador robusto.

Sea entonces $E(w) E^T(w) = E$, y recordemos que se ha definido

$$\| e \|_S^2 = e^T S(\theta) e.$$

$$e^T S_\infty(\theta) E S_\infty(\theta) e \leq \| S_\infty(\theta) e \|^2 \| E \| \quad (6.41)$$

Por otra parte, puede escribirse a e como combinación lineal de los autovectores de $S_\infty(\theta)$:

$$\begin{aligned} e &= \sum_{i=1}^n e_i \mu_i \\ S_\infty(\theta) e &= \sum_{i=1}^n e_i \lambda_i \mu_i \end{aligned}$$

Siendo λ_i los autovalores de $S(\theta)$, μ_i sus autovectores y e_i las componentes de e .

Entonces

$$\| S_\infty(\theta) e \|^2 = \sum_{i=1}^n e_i^2 \lambda_i^2$$

Análogamente,

$$\| e \|_S^2 = \| S_\infty(\theta)^{1/2} e \|^2 = \sum_{i=1}^n e_i^2 \lambda_i \quad (6.42)$$

Debido a la forma de $S_\infty(\theta)$, para valores de $\theta \geq 1$ todos sus autovalores (que son positivos por ser una matriz positiva definida) son menores o iguales que 1, con lo cual se tiene

$$\sum_{i=1}^n e_i^2 \lambda_i^2 \leq \sum_{i=1}^n e_i^2 \lambda_i$$

En consecuencia

$$\| S_\infty(\theta) e \|^2 \leq \| e \|_S^2 \quad (6.43)$$

Reemplazando (6.43) en (6.41), se obtiene

$$\| S_\infty(\theta) e \|^2 \| E \|^2 \leq \| e \|_S^2 \| E \|^2$$

Por lo tanto, el último término de la expresión (6.40) puede escribirse

$$\begin{aligned} & \frac{1}{\gamma^2} e^T S_\infty(\theta) E(w) E^T(w) S_\infty(\theta) e \leq \\ & \leq \frac{1}{\gamma^2} \| E(w) E^T(w) \| \| S_\infty(\theta) e \|^2 \leq \frac{1}{\gamma^2} \| E(w) E^T(w) \| \| e \|_S^2 \end{aligned}$$

Obteniéndose, en (6.40)

$$\| e \|_S^2 \left(-\theta + \frac{2(S\delta)^{1/2}kn^2}{\theta} + \delta + \frac{1}{\gamma^2} \| E(w) E^T(w) \| \right) < 0$$

Finalmente, se llega a la siguiente condición

$$-\theta + \frac{2(S\delta)^{1/2}kn^2}{\theta} + \delta + \frac{1}{\gamma^2} \| E(w) E^T(w) \| < 0 \quad (6.44)$$

Esta última expresión conduce a un nuevo diseño del observador TFKE, aumentando su robustez frente a perturbaciones externas.

6.2.3 Resultados y simulaciones

Para rediseñar el observador, posibilitando la atenuación de perturbaciones a un nivel predeterminado dado por el valor de γ , es preciso en primer lugar acotar $E(w)$. El cálculo debe realizarse en función de las variables originales, ya que debido a la forma de la transformación $T(z)$ no es posible obtener en forma explícita al vector w .

Observemos que la norma de $E = E(w) E^T(w)$ está dada por el máximo autovalor de dicha matriz para cada terna z_1, z_2, z_3 . Para realizar esta acotación, graficamos $\|E\|$ en función de z_2 y z_3 para distintos valores de z_1 . Los resultados se muestran en el Apéndice E.

Luego de analizar las figuras que se muestran en dicho apéndice, se definió el valor de $\|E\| = \|E(w) E^T(w)\|$.

Por otra parte, en la expresión (6.44), se conocen los valores de S , k , y n (ver [54]). Además, se elige $\gamma = \frac{1}{10} \rightarrow \gamma^2 = \frac{1}{100}$. Con todo esto, la desigualdad (6.44), que permite obtener el valor de θ que robustiza al observador queda

$$-\theta^2 + (k_1 + \delta)\theta + k_2 \delta^{1/2} < 0 \quad (6.45)$$

donde k_1 y k_2 dependen de S , k , n y de la cota obtenida.

Nótese que para definir el valor de δ es preciso hallar una cota del mínimo autovalor de $S_\infty(\theta)$. Una forma de calcularlo se presenta en [63]; otra posibilidad es realizar el cálculo en forma iterativa. Sin embargo, en este caso y debido a las dimensiones relativas de la cota de $E(w)$ y $\frac{1}{\gamma^2}$, el valor de δ no influye prácticamente en la elección de θ .

Finalmente se obtiene un valor de θ , que proporciona una gran velocidad de convergencia del observador, pero también un transitorio que puede tener sobrepicos muy elevados y abruptos. Si redujéramos la exigencia en cuanto al nivel de atenuación de perturbaciones (por ejemplo, $\gamma^2 = \frac{1}{10}$), el valor de θ requerido disminuye en el mismo orden.

Es decir, para lograr convergencia hay un compromiso entre una mayor atenuación de perturbaciones (menor γ) y un buen comportamiento transitorio y baja amplificación del ruido (bajo θ).

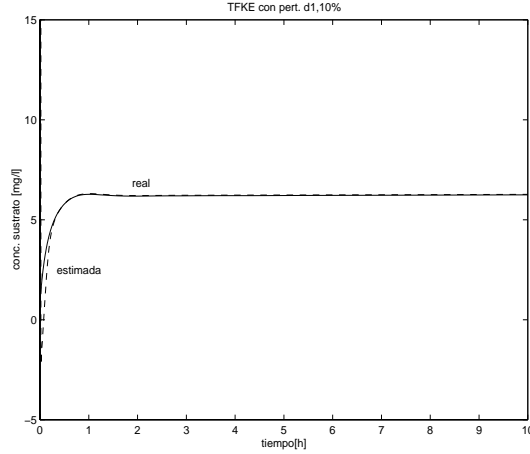


Figura 6.5: Sistema con perturbación en el caudal de entrada, 10% del valor nominal, conc. sustrato

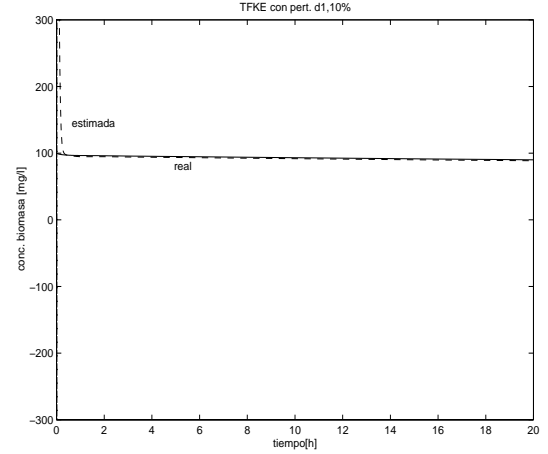


Figura 6.6: Sistema con perturbación en el caudal de entrada, 10% del valor nominal, conc. biomasa

6.2.3.1 Simulaciones para el sistema con perturbaciones

Se realizaron simulaciones con el nuevo valor de θ para distintos niveles de las perturbaciones en el caudal de alimentación y en la concentración de sustrato. Los resultados se presentan en las figuras 6.5 a 6.17.

En las Figuras 6.5 a 6.10 se muestra el desempeño del observador para la estimación de la concentración de sustrato y biomasa cuando se produce un aumento del caudal de la alimentación a la planta; para (6.5) a (6.7) del orden del 10 % y del 50 % para (6.8) a (6.10).

Las Figuras 6.11 a 6.15 presentan la performance del observador cuando se estiman las concentraciones de sustrato y biomasa, respectivamente, para un aumento en la concentración de sustrato en la alimentación del orden del 20 % y del 50 %.

Finalmente, se muestra la evolución de la concentración de sustrato (real y estimada) en el caso en que aparecen simultáneamente las dos perturbaciones: una variación del 20 % en el caudal de entrada, y del 20 % en la concentración de sustrato en la alimentación (Figuras 6.16 y 6.17).

Como conclusión, observamos que el observador TFKE rediseñado pa-

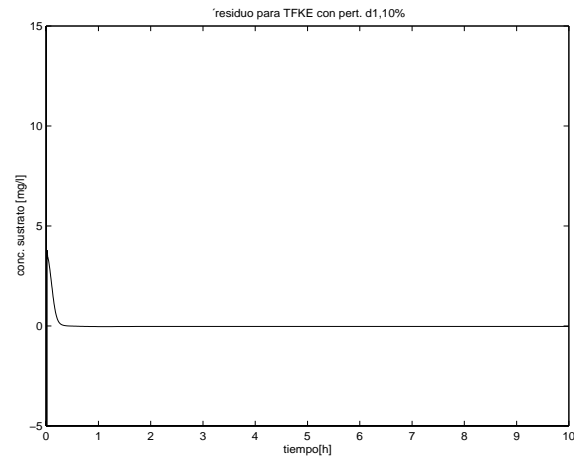


Figura 6.7: Sistema con perturbación en el caudal de entrada, 10% del valor nominal, residuo

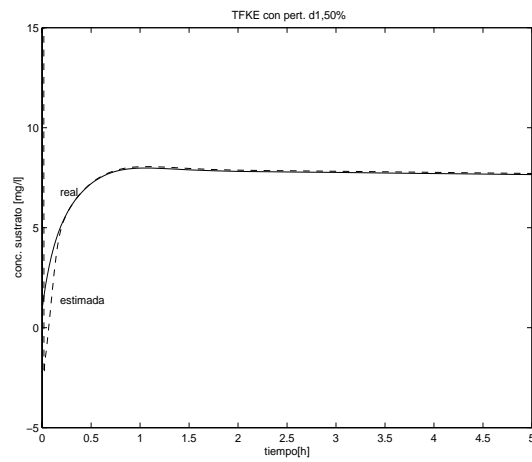


Figura 6.8: Sistema con perturbación en el caudal de entrada, 50% del valor nominal, conc. sustrato

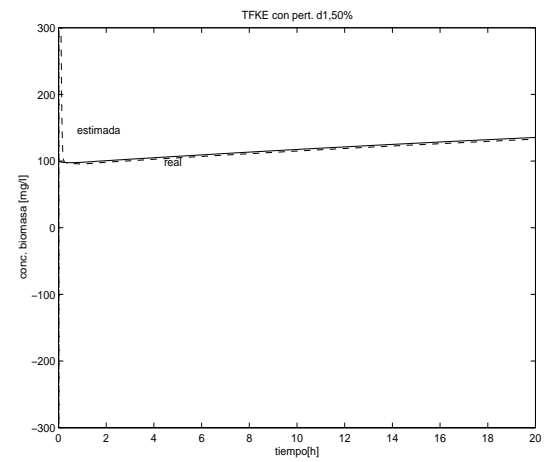


Figura 6.9: Sistema con perturbación en el caudal de entrada, 50% del valor nominal, conc. biomasa

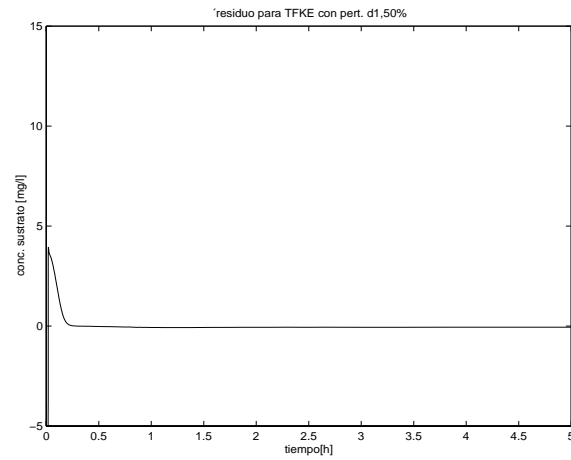


Figura 6.10: Sistema con perturbación en el caudal de entrada, 50% del valor nominal, residuo

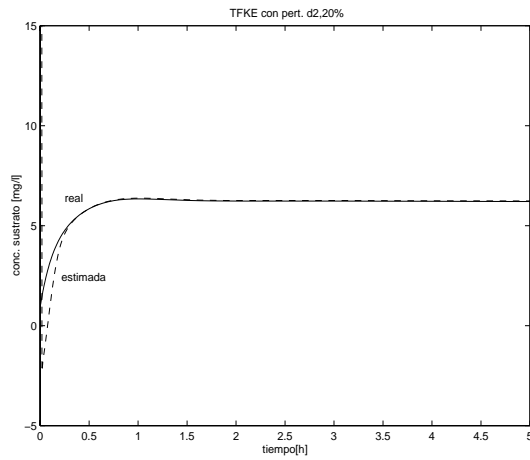


Figura 6.11: Sistema con perturbación en la conc.de sustrato, 20% del valor nominal, conc. sustrato

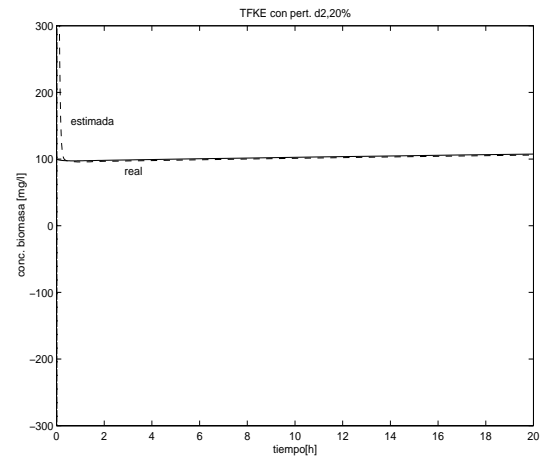


Figura 6.12: Sistema con perturbación en la conc. de sustrato, 20% del valor nominal, conc. biomasa

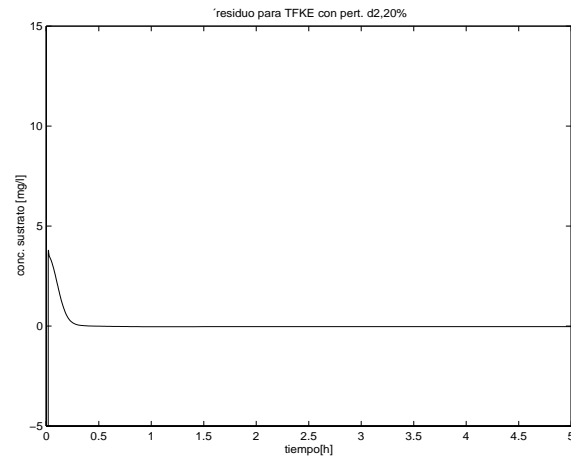


Figura 6.13: Sistema con perturbación en la conc. de sustrato, 20% del valor nominal, residuo

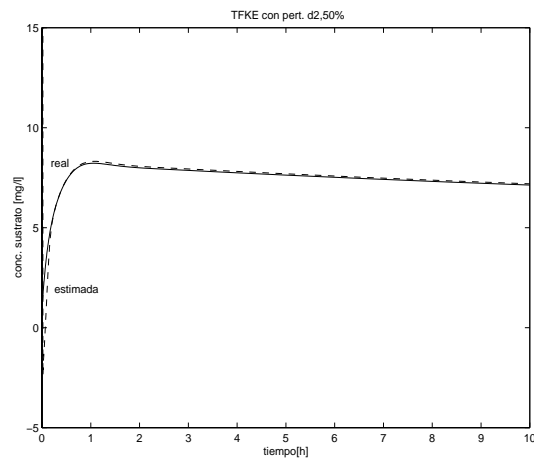


Figura 6.14: Sistema con perturbación en la conc.de sustrato, 50% del valor nominal, conc. sustrato

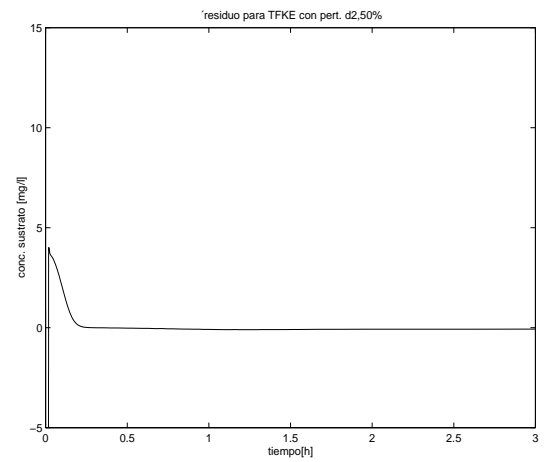


Figura 6.15: Sistema con perturbación en la conc. de sustrato, 50% del valor nominal, residuo.

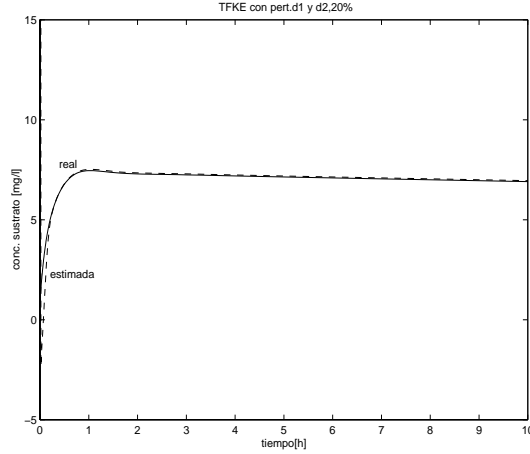


Figura 6.16: Sistema con ambas perturbaciones, 20% del valor nominal, conc. sustrato

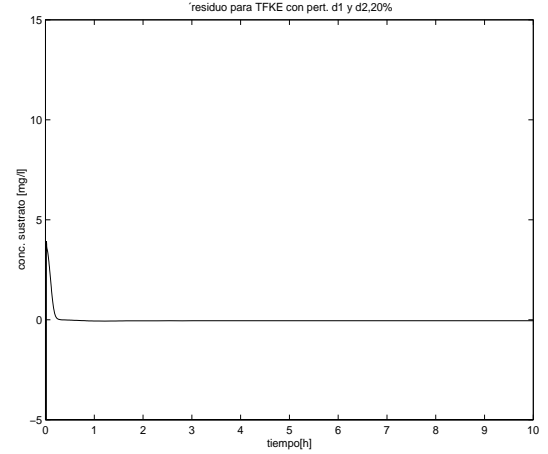


Figura 6.17: Sistema con ambas perturbaciones, 20% del valor nominal, residuo

ra lograr la atenuación de perturbaciones presenta un muy buen desempeño tanto en los casos en que se producen en forma aislada las dos perturbaciones como en aquél en que se manifiestan simultáneamente. En los gráficos que muestran los residuos puede verse que el transitorio inicial, si bien presenta un sobrepico considerable -lo cual es esperable, debido a las consideraciones realizadas previamente respecto de los valores elevados de θ , la ganancia del observador- tiene una duración breve. En todos los casos puede notarse que el residuo es prácticamente nulo luego de la primera media hora de iniciada la perturbación, lo que indica un muy buen comportamiento del generador de residuos para este sistema.

6.3 Análisis comparativo de los observadores propuestos

En este capítulo hemos propuesto tres observadores para la planta de tratamiento de aguas residuales, cuyo objetivo final es la detección de fallas. El primero, *Observador de orden reducido*, puede implementarse de manera sencilla con las técnicas usuales para los observadores de tipo identidad. Si bien funciona correctamente para los valores nominales de los parámetros, una pequeña desviación de los mismos lleva al observador fuera de la región

de convergencia. Dicho de otro modo, su robustez es prácticamente nula, lo que lo torna inviable para un sistema como el que estudiamos, en el cual es esperable un cierto rango de incertidumbre en los parámetros del modelo.

El segundo, *Observador de estructura variable*, posee como ya dijimos cierta robustez inherente al no requerir el conocimiento exacto de las no linealidades del sistema. Sin embargo, su comportamiento frente a perturbaciones externas, aun pequeñas, es inadecuado.

El tercero, *Observador del tipo Filtro de Kalman extendido*, presenta un correcto desempeño en un entorno más amplio que los precedentes. En cuanto a su velocidad de convergencia, es mucho mayor que la del OEV, aun para valores de θ pequeños. El estudio de la atenuación de perturbaciones permitió lograr un buen comportamiento, incluso en el caso en que se presentan variaciones en el caudal de la alimentación y en la concentración de sustrato de entrada del orden del 50 %. Si bien el empleo de observadores de alta ganancia como éste se ve limitado por la inevitable amplificación del ruido (ver [26]), cabe destacar que nuestro planteo para atenuar perturbaciones puede utilizarse de la misma manera para la atenuación del ruido del proceso, aunque esta importante perspectiva no será tratada en esta tesis.

6.4 Detección de fallas en la planta

Una vez logrado el objetivo de generar residuos robustos mediante la atenuación de las perturbaciones del sistema, mostraremos cómo se reflejan las fallas en el error de observación. Dicho de otro modo, comprobaremos que el residuo posee la propiedad de mínima sensibilidad frente a las perturbaciones y una adecuada sensibilidad a las fallas.

6.4.1 El modelo con fallas

En la Sección 5.5, se incluyeron las perturbaciones en el modelo matemático de la planta. Siguiendo un razonamiento análogo pueden modelarse las fallas, con lo cual una descripción del sistema que las incluya quedaría expresada por

$$\dot{z} = L(z, u, d, f)$$

Del mismo modo que las perturbaciones, las fallas pueden aproximarse expandiendo $L(z, u, d, f)$ en serie de Taylor alrededor de los valores nominales de los parámetros, desestimando los términos de segundo y mayor orden.

Consideraremos en este caso las siguientes fallas:

1. Falla 1 (f_1): Disminución de la concentración de biomasa por aparición de una sustancia tóxica para los microorganismos.
2. Falla 2 (f_2): Disminución del caudal de reflujo por bloqueo de una válvula.

Falla 1.- La concentración de lodo biológico (biomasa) aumenta al principio del proceso, durante el período en que una concentración sustancial de sustrato se encuentra presente para proporcionar alimentación abundante para mantener el crecimiento de los microorganismos. Cuando la concentración de sustrato disminuye considerablemente y no hay suficiente alimento para mantener el crecimiento de los microorganismos, éstos comienzan a consumirse entre sí y desciende la concentración de biomasa, ya que la velocidad de destrucción de las células excede la de síntesis. Esta es la fase llamada de “respiración endógena”.

El parámetro de diseño b_H se define como la fracción de lodo biológico por unidad de tiempo oxidada durante el proceso de respiración endógena, es decir, la utilización de energía en reacciones para mantenimiento celular. Pero además involucra otros factores, como la muerte de células microbianas por predación debida a algún agente externo. Por ejemplo, $b_H = 0.1 \text{ 1/d}$ indica que el 10% de la cantidad total de biomasa presente en el biorreactor, en cualquier momento, se oxida por día. El factor b_H se denomina coeficiente de decaimiento de biomasa o coeficiente de descomposición microbiana.

En consecuencia, una variación de la concentración de biomasa por la introducción, por ejemplo, de una sustancia tóxica para los microorganismos se vería reflejada en este parámetro. Por lo tanto, tomemos a b_{H0} como su valor nominal y consideremos ahora

$$b_H = b_{H0} + \Delta b$$

De acuerdo a las ecuaciones del modelo, esta falla se introduce al sistema

mediante

$$\frac{\partial L(z, u, d, f)}{\partial b_H} = \begin{bmatrix} 0 \\ (1 - f_p)(z_3 + x_3^*) \\ -(z_3 + x_3^*) \end{bmatrix} \quad (6.46)$$

Falla 2.- Si se produjera el bloqueo parcial de la válvula en la corriente de reflujo, el caudal del mismo se vería afectado y en consecuencia disminuiría la cantidad de biomasa activa que se realimenta al biorreactor para mantener la reacción. Si llamamos q_{R0} al valor nominal del caudal de reflujo,

$$q_R = q_{R0} + \Delta q_R$$

Considerando las ecuaciones del modelo, esta falla se introduce al sistema mediante

$$\frac{\partial L(z, u, d, f)}{\partial q_R} = \begin{bmatrix} \frac{-(z_1 + x_1^*)}{V} \\ 0 \\ \frac{-q_w}{V}(z_3 + x_3^*) \frac{q_w - q_{F0}}{(q_w + q_{R0})^2} \end{bmatrix} \quad (6.47)$$

6.4.1.1 Matriz de distribución de fallas

En síntesis y de acuerdo a las expresiones obtenidas para ambas fallas (f_1 y f_2), la matriz de distribución quedaría dada por

$$R(z) = \begin{bmatrix} \frac{\partial L(z, u, b_H, q_R)}{\partial b_H} & \frac{\partial L(z, u, b_H, q_R)}{\partial q_R} \end{bmatrix} \quad (6.48)$$

El término que modeliza las fallas en el sistema es

$$\begin{aligned} R(z)f &= \begin{bmatrix} 0 & \frac{-(z_1 + x_1^*)}{V} \\ (1 - f_p)(z_3 + x_3^*) & 0 \\ -(z_3 + x_3^*) & \frac{-q_w}{V}(z_3 + x_3^*) \frac{q_w - q_{F0}}{(q_w + q_{R0})^2} \end{bmatrix} \begin{bmatrix} \Delta b_H \\ \Delta q_R \end{bmatrix} \\ &= \begin{bmatrix} R_1 & R_2 \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \end{aligned} \quad (6.49)$$

Entonces, la expresión completa del modelo incluyendo fallas y perturbaciones queda como sigue.

$$\begin{aligned}
\dot{z} &= A^* z + \begin{bmatrix} e \\ e-1 \\ 1 \end{bmatrix} \mu_R(z_1, z_2, z_3) + \\
&\quad + E(z) \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} + R(z) \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} + \begin{bmatrix} K_a(S_{O,sat} - z_1 - x_1^*) \\ 0 \\ 0 \end{bmatrix} u \\
y &= \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} z
\end{aligned} \tag{6.50}$$

donde A^* es la matriz de la ecuación (5.10), $\mu_R(z_1, z_2, z_3)$ y $E(z)$ fueron definidos en las Secciones (5.4.7) y (5.5) respectivamente.

6.4.2 Evaluación del residuo

El segundo paso en la DEF es la evaluación de los residuos, proceso de toma de decisión que consiste en establecer un umbral lógico mediante una función de decisión. En los problemas prácticos, en los cuales no siempre se puede lograr el desacoplamiento total de las perturbaciones y las mediciones se ven afectadas por ruido deben asignarse, como ya se dijo previamente, umbrales diferentes de cero.

En estos casos, la evaluación robusta del residuo es la única forma de lograr una baja frecuencia de falsas alarmas con una aceptable sensibilidad a las fallas.

En primer lugar debe elegirse una función de evaluación y, basándose en ella, determinar el umbral correspondiente. En la práctica se usa frecuentemente la raíz media cuadrática (*rms*) [23].

$$\| r(t) \|_e = J(\tau) = \left(\tau^{-1} \int_0^\tau r^T(t) r(t) dt \right)^{1/2}$$

donde τ denota la ventana de evaluación en el dominio del tiempo y $\| \cdot \|_e$ indica función de evaluación.

En la práctica, la función de evaluación del residuo puede ser seleccionada teniendo en cuenta cómo pueden evaluarse las fallas a ser detectadas y qué clase de información se posee acerca del proceso y posibles fallas.

Una vez seleccionada la función de evaluación se determina el umbral, con el objetivo de reducir o prevenir falsas alarmas pero sin que esto produzca la pérdida o no detección de fallas reales. Para la elección del umbral debe tomarse en cuenta el ruido en la medición, de modo de garantizar que el disparo de alarma se produzca cuando la señal haya superado un cierto valor, definido por consideraciones probabilísticas.

Entonces, en ausencia de fallas $\| r \|_e$ debería ser menor que el umbral y por lo tanto

$$J = \sup_{f=0} \| r \|_e$$

Como se dijera previamente, el residuo robusto puede ser utilizado para detectar fallas mediante

$$\| r(t) \| = \begin{cases} < J, & \text{si no hay fallas} \\ \geq J, & \text{si hay fallas} \end{cases}$$

siendo J el umbral.

Se han desarrollado variadas técnicas para la selección de umbrales, entre las cuales se destaca, por ejemplo, el umbral adaptivo [23]. Dado que en esta tesis no nos hemos centrado en el análisis de los ruidos de medición soslayaremos esta perspectiva, que puede ser estudiada en trabajos futuros.

6.4.3 Simulaciones y resultados para el proceso con fallas

A continuación, se presentan las simulaciones que ilustran el comportamiento del sistema cuando se producen en él las fallas f_1 y f_2 . En las figuras que siguen, se observa la evolución del residuo a lo largo de 20 horas de proceso. En las primeras 10 horas el sistema no presenta fallas ($r = 0$), mientras que a $t = 10h$ se produce ya sea una variación en el caudal de reflujo q_R (f_2) como en el coeficiente de decaimiento b_H (f_1).

Falla 1: En el caso de la variación del coeficiente de decaimiento, es decir, la muerte de microorganismos por efecto de un agente externo al proceso (por ejemplo, una sustancia tóxica), consideramos dos posibilidades:

- A) falla abrupta (tipo escalón): se produciría si repentinamente la alimentación a la planta ingresara con una elevada concentración

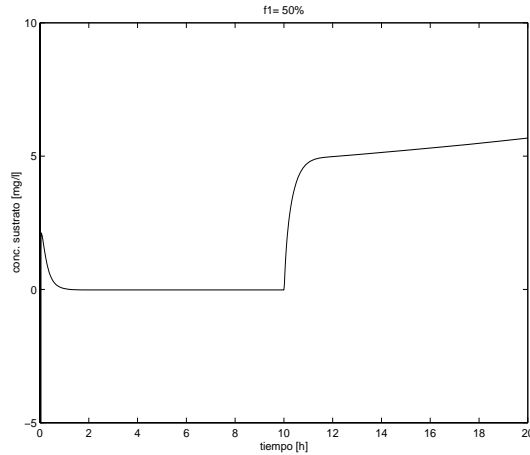


Figura 6.18: Falla 1. Variación de b_H en un 50 % respecto del valor nominal.

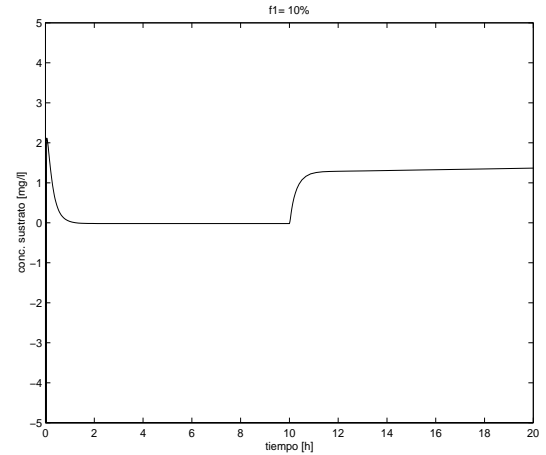


Figura 6.19: Falla 1. Variación de b_H en un 10 % respecto del valor nominal.

de una sustancia tóxica. El efecto sobre el coeficiente de descomposición microbiana sería igualmente abrupto, ya que ocasionaría la muerte de gran cantidad de microorganismos. Las figuras 6.18 a 6.20 muestran la aparición de una variación de tipo escalón en b_H del orden del 50 %, del 10 % y del 5 %, respectivamente, respecto de su valor nominal.

Si se observan los datos proporcionados por el generador de residuos, para el caso en que la variación de b_H es del 50 % respecto de su valor nominal la falla podría detectarse a los pocos minutos de sucedida, lo que indica un buen desempeño del generador. Para variaciones porcentuales menores, como es esperable, el efecto sobre el residuo también disminuye, aunque una falla del 5 % es detectable con un retardo de una hora aproximadamente.

- B) falla suave o incipiente (tipo rampa): se produciría si la alimentación a la planta contuviera una concentración inicialmente pequeña de la sustancia tóxica, o si el agente externo actuara lentamente sobre el metabolismo de los microorganismos. En este caso la detección temprana de la falla puede evitar el lavado del biorreactor, que produciría una salida fuera de especificación durante un tiempo estrechamente relacionado con el retardo con que se detecta la falla. Se realizaron simulaciones para dos valores di-

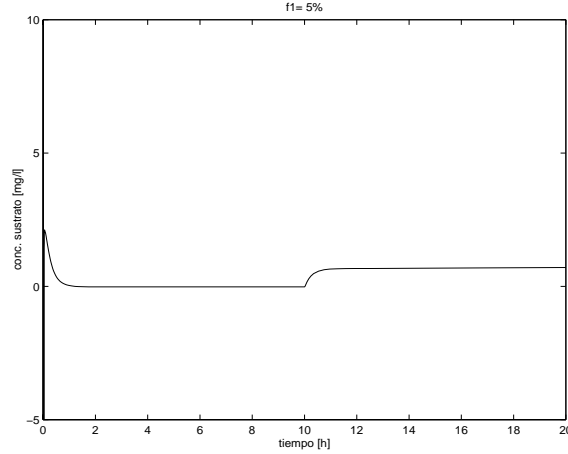


Figura 6.20: Falla 1. Variación de b_H en un 5 % respecto del valor nominal.

ferentes de pendiente de la rampa: 25 % y 2.5 % del valor nominal de b_H por hora, respectivamente. En las figuras se observa que el generador de residuos manifiesta una situación anormal, observable aproximadamente a los 20 minutos de su iniciación en el primer caso, mientras que en el segundo el retardo es de menos de una hora, lo que constituye un excelente desempeño teniendo en cuenta la pequeña pendiente de la rampa.

Notemos que el parámetro b_H tiene una cota superior natural, ya que representa una fracción de decaimiento. Sin embargo, el sistema de detección de fallas debería producir un disparo de alarma antes de que se alcance dicho valor máximo.

Falla 2: En este caso, analizamos los datos proporcionados por el generador de residuos cuando se produce un bloqueo en la válvula reguladora del reflujo.

En las figuras 6.25 a 6.29 puede observarse el comportamiento del sistema cuando se produce una variación en el caudal de reflujo del orden del 50 % y del 10 %, respectivamente, respecto de su valor nominal. Notamos que en ambos casos la falla es detectable a los pocos minutos de sucedida, con lo cual puede actuarse rápidamente para evitar una degradación considerable en la performance del sistema.

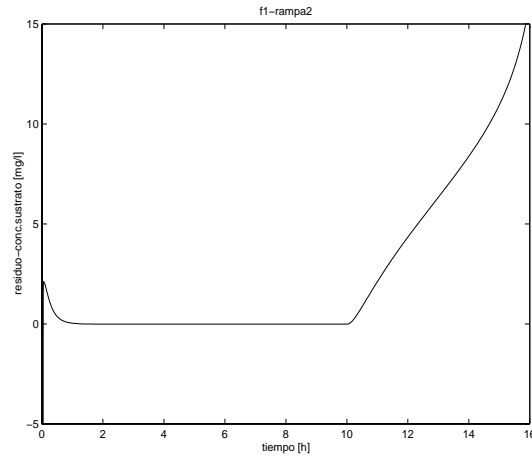


Figura 6.21: Falla 1. Variación de b_H según una rampa de pend. 25% del valor nominal.

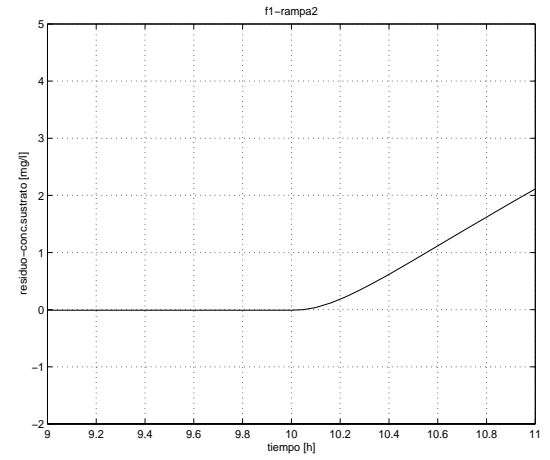


Figura 6.22: Falla 1. Variación de b_H según una rampa, $9 h \leq t \leq 11 h$.

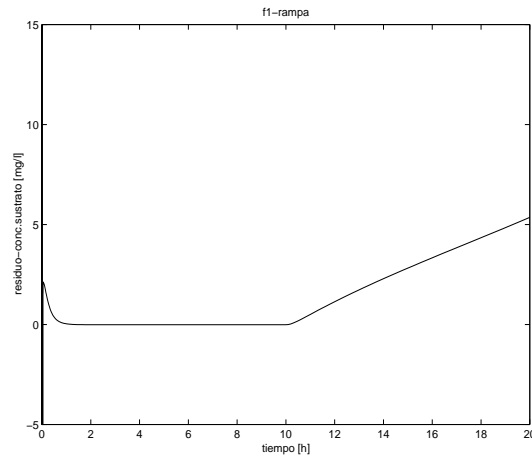


Figura 6.23: Falla 1. Variación de b_H según una rampa de pend. 2.5% del valor nominal.

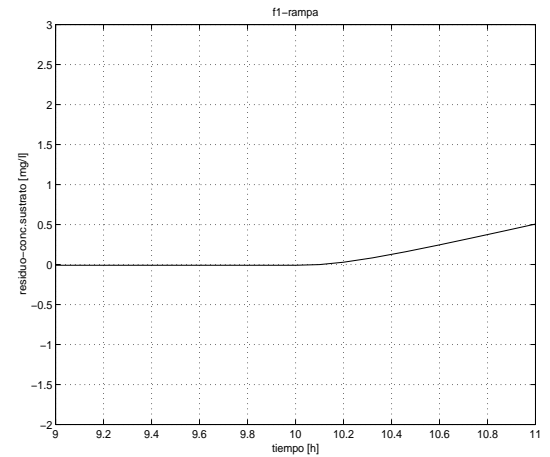


Figura 6.24: Falla 1. Variación de b_H según una rampa, $9 h \leq t \leq 11 h$.

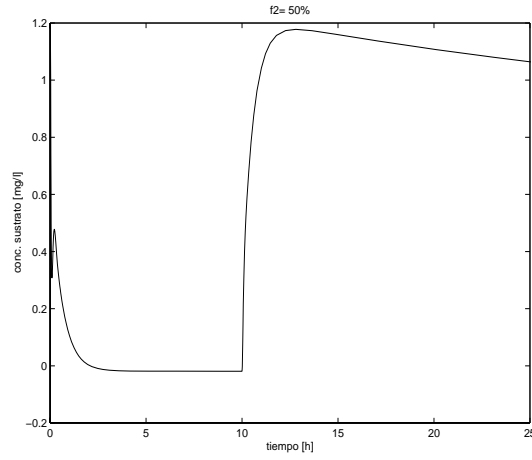


Figura 6.25: Falla 2. Variación de q_R en un 50 % respecto del valor nominal.

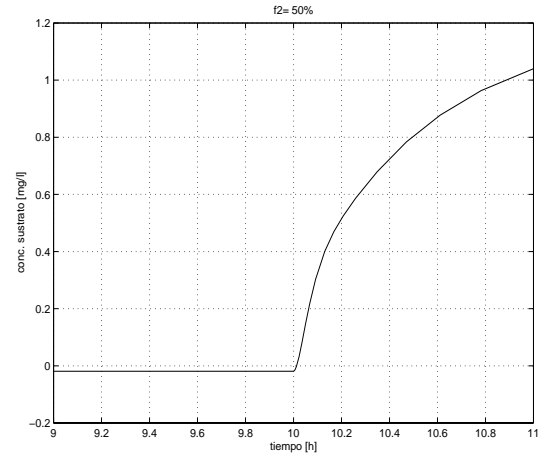


Figura 6.26: Falla 2. Variación de q_R en un 50 % respecto del valor nominal $9 h \leq t \leq 11 h$.

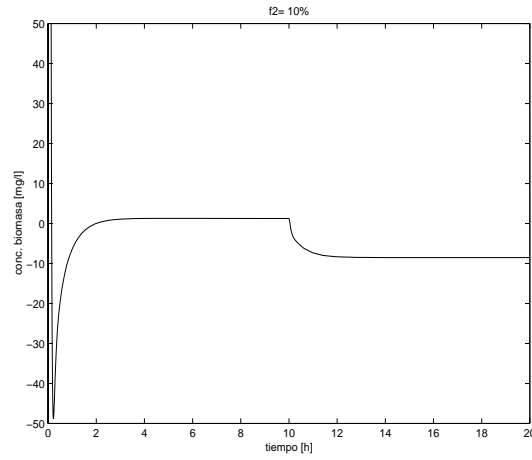


Figura 6.27: Falla 2. Variación de q_R en un 10 % respecto del valor nominal, conc.biomasa.

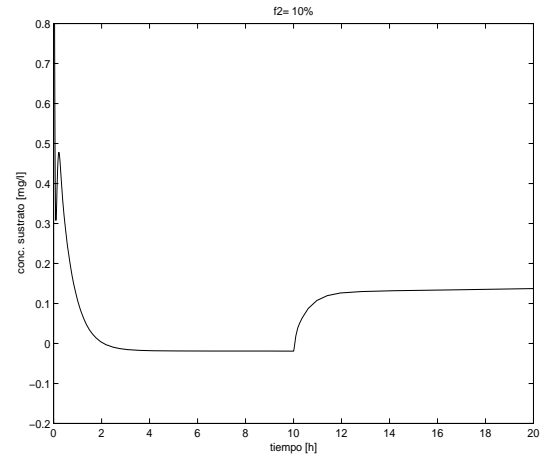


Figura 6.28: Falla 2. Variación de q_R en un 10 % respecto del valor nominal.

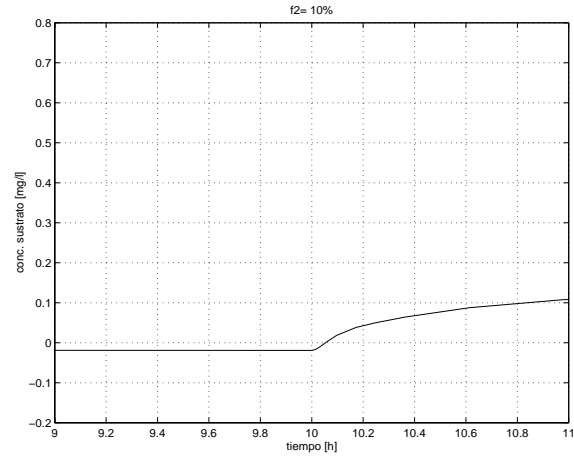


Figura 6.29: Falla 2. Variación de q_R en un 10 % respecto del valor nominal, $9 h \leq t \leq 11 h$.

De las figuras anteriores puede concluirse que el generador de residuos presenta una muy buena performance para la detección de estas fallas en la planta. En futuras investigaciones, podrían incluirse otras fallas en el modelo, así como avanzar en su aislamiento e identificación.

Capítulo 7

Conclusiones

En este trabajo hemos estudiado la detección de fallas en sistemas de control automático, y su aplicación a una planta de tratamiento de efluentes.

Presentaremos a continuación nuestras conclusiones, luego desarrollaremos una breve síntesis de los temas tratados a lo largo de la tesis así como de las principales contribuciones, y por último mencionaremos las posibles extensiones del trabajo.

7.1 Conclusiones

En la actualidad y debido a la creciente complejidad de los procesos y sistemas de control, se incrementa la necesidad de proveer a los mismos de un alto grado de confiabilidad. Para ello, es útil contar con sistemas que permitan detectar precozmente posibles fallas del proceso, informando mediante un disparo de alarma la ocurrencia de una situación anormal en la planta. Focalizando nuestra atención en la detección y aislamiento de fallas en sistemas no lineales de control automático, hallamos un vasto campo de estudio, con numerosos aspectos aún por analizar.

En particular, observamos que las técnicas de diseño de observadores no lineales para generación de residuos pueden no ser aplicables, en la práctica, a determinados sistemas. Teniendo en cuenta que la observabilidad en los SNL no es una propiedad global, sino que por el contrario depende del punto de operación, y considerando el fundamental requerimiento de robustez que plantea la DEF, se evidencia que no todos los ONL satisfacen las condiciones

para desempeñarse adecuadamente como generadores de residuos. Para el sistema de tratamiento de efluentes, hemos diseñado tres observadores (*de orden reducido, de estructura variable y tipo filtro de Kalman extendido*), entre los cuales elegimos el TFKE por presentar la mejor performance frente a perturbaciones externas.

Como ya se dijo, la eficiencia de la DEF está estrechamente relacionada con la posibilidad de rechazar o al menos atenuar las perturbaciones que pueden aparecer en la planta. En rigor, una propiedad fundamental de un buen generador de residuos consiste en poseer una mínima sensibilidad frente a perturbaciones junto con una máxima sensibilidad a las fallas. Las características particulares de los sistemas de tratamiento de aguas residuales, algunos de cuyos parámetros presentan fluctuaciones inevitables por causas externas al sistema, ponen de manifiesto que este análisis no puede soslayarse. En una primera aproximación, observamos que las condiciones para el desacoplamiento total de perturbaciones son altamente restrictivas y no se cumplen en nuestro caso. Por ello, sólo podemos garantizar su atenuación a un nivel prefijado. Utilizamos una técnica similar a la desarrollada en [33] para diseñar controladores estabilizantes globales con atenuación de perturbaciones. Nuestro aporte consistió en plantear un nuevo problema, el de atenuación de perturbaciones para observadores no lineales (lo que en sistemas lineales sería el dual del anterior), extendiendo además su aplicación a un campo vectorial multivariable. Los resultados permitieron un rediseño del observador TFKE, lográndose un muy buen desempeño para variaciones considerables de los parámetros con respecto a su valor nominal.

Luego de desarrollar el modelo matemático de la planta a partir de balances de masa, así como de ecuaciones cinéticas que reflejan el comportamiento de los microorganismos, incluimos en él tanto las perturbaciones como las fallas que más comúnmente pueden afectar al proceso. Es decir, construimos un modelo apropiado en función de nuestro objetivo y aplicamos en él las herramientas previamente obtenidas, tanto para atenuar perturbaciones como para detectar fallas. Los resultados muestran que nuestro observador de alta ganancia, rediseñado para atenuar perturbaciones, pudo implementarse con muy buen desempeño para la detección de fallas en la planta de tratamiento de efluentes.

7.2 Breve síntesis

Detección y aislamiento de fallas: Desarrollamos aquí las técnicas más usuales para DEF y generación de residuos, haciendo énfasis en aquéllas que se basan en el uso de observadores. En particular, resulta relevante para nuestro trabajo el estudio del diseño de los Observadores de entrada desconocida (OED), cuyo principio consiste en lograr que el vector de estimación no se vea afectado por las perturbaciones.

El requerimiento de robustez es especialmente restrictivo en los sistemas no lineales, para los cuales mostramos diferentes métodos de generación de residuos mediante ONL y estudiamos la posibilidad del desacoplamiento total de las perturbaciones que pueden afectar al sistema.

Desacoplamiento y atenuación de perturbaciones: Para que el sistema de DEF sea eficiente, los efectos de las perturbaciones no deberían reflejarse en el residuo. En los sistemas reales esto no puede lograrse fácilmente, y menos aún si responden a un modelo no lineal. Cuando no se verifican las condiciones necesarias para el desacoplamiento total, puede estudiarse la posibilidad de un desacoplamiento aproximado, optimizando índices de desempeño que midan la relación entre la sensibilidad a las fallas y la sensibilidad frente a perturbaciones. En los sistemas no lineales lograr que la salida sea completamente independiente de las entradas desconocidas impone restricciones severas al sistema, por lo cual sólo es posible en casos especiales. En consecuencia es útil proponer un nuevo camino que posibilite la atenuación de las perturbaciones a un nivel prefijado. La reducción del efecto de las entradas desconocidas en la salida del sistema - que en este caso es el error de observación- puede medirse en términos de la *ganancia* L_2 , concepto que juntamente con el teorema de Artstein-Sontag permite plantear el problema en función de una desigualdad de Hamilton-Jacobi. Hemos aplicado los resultados obtenidos por Isidori para controladores a los observadores no lineales, estudiando la atenuación de perturbaciones en el modelo de error. Es decir, desarrollamos lo que en sistemas lineales se conoce como el dual del problema planteado en [33]. Además realizamos la extensión de dichos resultados a un campo vectorial multivariable.

Sistemas de tratamiento de efluentes: Presentamos las características generales de los procesos de tratamiento biológico de aguas residuales,

en especial el método de lodos activados. Desarrollamos los balances de materia y ecuaciones cinéticas que permiten construir el modelo matemático -no lineal- de la planta, incluyendo en él en una segunda etapa el modelado de perturbaciones. Por último, describimos algunas de las fallas típicas que suelen presentarse en el proceso, a fin de construir un modelo de fallas que permita luego instrumentar un sistema para su detección y aislamiento.

Detección de fallas en la planta: Estudiamos la aplicación de la DEF mediante observadores no lineales a una planta de tratamiento de aguas residuales. Desarrollamos en primer lugar el diseño de tres observadores propuestos para el sistema: *observador de orden reducido*, *observador de estructura variable*, *observador del tipo Filtro de Kalman extendido*. Realizamos las simulaciones que ilustran su desempeño, estudiando luego su robustez frente a perturbaciones externas. De la comparación entre ellos puede concluirse que el observador TF-KE presenta la mejor performance en cuanto al rechazo de perturbaciones. Sin embargo, dado que no se cumplen en el sistema las condiciones para el desacoplamiento total, debe replantearse el problema en términos de la atenuación de perturbaciones a un nivel predefinido. Intodujimos en el modelo dos perturbaciones externas que aparecen frecuentemente en la planta: variación en la concentración de sustrato de entrada y variación en el caudal de alimentación. El rediseño del observador para un nivel predefinido de atenuación de las perturbaciones se muestra efectivo, incrementando la robustez del mismo al mostrar un desempeño adecuado para variaciones de los parámetros del orden del 50 % de su valor nominal. Finalmente, estudiamos dos fallas que pueden aparecer en el sistema: disminución de la concentración de biomasa en el biorreactor por aparición de una sustancia tóxica y disminución del caudal de reflujo por bloqueo de la válvula reguladora. Observamos que el generador de residuos se desempeña correctamente en ambos casos, permitiendo detectar las fallas en un tiempo razonable, de modo de evitar efectos graves sobre el proceso.

En síntesis, las dos mayores contribuciones de esta tesis son

- i) el esquema de atenuación de perturbaciones.

- ii) la obtención de las condiciones para lograr convergencia del generador de residuos.

7.3 Trabajo futuro

Del estudio de las distintas técnicas para ONL se observa que los desarrollos teóricos no siempre pueden aplicarse en forma directa a los procesos reales. En particular, en el caso del observador TFKE la determinación del parámetro θ se realizó conservativamente, debido al cálculo de la cota para la transformada de la perturbación. Una posible extensión del trabajo consiste en ajustar el cálculo de dicha cota. Esto a su vez posibilitaría la obtención de un valor de θ menos conservativo, lo cual es importante dado que existe un compromiso entre la atenuación de perturbaciones y un buen comportamiento transitorio con baja amplificación del ruido. Con respecto a esto último, cabe destacar que nuestra propuesta para atenuación de perturbaciones podría aplicarse, análogamente, para la atenuación del ruido del proceso, lo que también plantea una interesante línea de trabajo futuro.

Otro aspecto en el cual se puede indagar es la evaluación robusta del residuo, así como la selección de umbrales de detección, en especial técnicas para el diseño de umbrales adaptivos que aumenten la eficiencia del sistema de DEF.

Por otra parte, hay un amplio campo para la investigación en el aislamiento e identificación de fallas en sistemas no lineales, ya que si bien este tema ha sido encarado por los investigadores, aún dista de alcanzar el desarrollo que presenta para los sistemas lineales. En particular, para los procesos de tratamiento de efluentes, resultaría de interés avanzar tanto en la modelización de nuevas fallas como en su aislamiento e identificación con el objetivo de, una vez realizada la detección, determinar la ubicación de las fallas (sensor, actuador, componente) y estimar su tamaño y naturaleza.

Apéndice A

Lema de Meyer-Kalman-Yacubovich

Teorema 1 *Consideremos el sistema*

$$\dot{x} = Ax + \Gamma^T(x, z, t)z \quad x \in \mathbb{R}^n, x(t_0) = x_0 \quad (\text{A.1})$$

$$\dot{z} = -\Lambda \Gamma(x, z, t)Px \quad z \in \mathbb{R}^p, z(t_0) = z_0 \quad (\text{A.2})$$

donde: $A_{n \times n}$ es una matriz Hurwitz; $\Gamma(x, z, t)_{p \times n}$ es una matriz de funciones suaves uniformemente acotadas para cada (x, z) acotado; $\Lambda_{p \times p}$, una matriz simétrica, definida positiva; $P_{n \times n}$ es la solución simétrica, positiva definida de la ecuación de Lyapunov

$$A^T P + P A = -Q$$

con Q una matriz simétrica, definida positiva.

Entonces:

- a) el punto de equilibrio $(x, z) = 0$ es uniformemente estable.
- b) las soluciones $x(t)$ y $z(t)$ de (A.1) son uniformemente acotadas $\forall t \geq t_0, \forall x_0 \in \mathbb{R}^n, \forall z_0 \in \mathbb{R}^p$.
- c) $\lim_{t \rightarrow \infty} \|x(t)\| = 0, \forall x_0 \in \mathbb{R}^n, \forall z_0 \in \mathbb{R}^p$.

Si además $\frac{\partial \Gamma}{\partial t}$ es uniformemente acotada para todo (x, z) acotado y existe una función α de clase K tal que:

$$z^T \Gamma(0, z, t) \Gamma^T(0, z, t) z \geq \alpha(\|z\|), \quad \forall t \geq t_0, \forall z \in \mathbb{R}^p,$$

entonces

d) el origen $(x, z) = 0$ es un equilibrio global, uniforme y asintóticamente estable para el sistema (A.1).

Lema (Meyer-Kalman-Yacubovich) 1 Sea $A_{n \times n}$ una matriz Hurwitz, y b, c^T , dos vectores reales de $n \times 1$. Si y sólo si el triplete (A, b, c) satisface la condición de ser estrictamente real positivo:

$$\operatorname{Re}\{c(j\omega I - A)^{-1}b\} > 0, \quad \forall \omega \in (-\infty, +\infty)$$

entonces, para una matriz $Q_{n \times n}$ simétrica, positiva definida, existe una matriz simétrica, positiva definida $P_{n \times n}$, un vector real $q_{n \times 1}$ y un número real positivo ϵ tal que:

$$A^T P + P A = -q q^T - \epsilon Q \quad (\text{A.3})$$

$$P b = c^T \quad (\text{A.4})$$

Apéndice B

Lema generalizado de Gronwall

Sean

- $\phi(t), \psi(t)$ reales y continuas, no negativas en $[a, b]$
- $k > 0$ una constante

Si para t en $[a, b]$,

$$\phi(t) \leq \psi(t) + k \int_a^t \phi(s) ds$$

entonces

$$\phi(t) \leq \psi(t) + \int_a^t e^{k(t-s)} \psi(s) ds$$

Nota:

Si $f : [0, \alpha] \rightarrow \mathbb{R}$ continua y no negativa, y se verifica:

$$f(t) \leq C + \int_0^t K f(s) ds, \quad \forall t \in [0, \alpha]$$

con $C \geq 0$ y $K \geq 0$, entonces

$$f(t) \leq C e^{Kt}, \quad \forall t \in [0, \alpha]$$

Apéndice C

Teorema de Artstein-Sontag

Consideremos un sistema descrito por ecuaciones de la forma:

$$\dot{x} = f(x) + g(x)u \quad (\text{C.1})$$

donde $f(x)$ y $g(x)$ son campos vectoriales suaves y $f(0) = 0$. Entonces existe una ley de realimentación casi suave $u = \alpha(x)$ que estabiliza global y asintóticamente el equilibrio $x = 0$ de (C.1) si y sólo si existe una función suave, definida positiva y propia $V(x)$ con las siguientes propiedades:

- $L_g V(x) = 0 \Rightarrow L_f V(x) < 0, \forall x \neq 0$.
- $\forall \epsilon > 0 \exists \delta > 0$ tal que si $x \neq 0$ satisface que $\|x\| < \delta$, entonces existe alguna u con $|u| < \epsilon$ tal que

$$L_f V(x) + L_g V(x)u < 0$$

Nota:

El teorema provee una fórmula simple para la ley de realimentación casi suave que estabiliza un sistema (C.1) global y asintóticamente. Esta es:

$$u = \alpha(x) = \begin{cases} 0 & \text{si } L_g V(x) = 0 \\ -\frac{L_f V(x) + \sqrt{[L_f V(x)]^2 + [L_g V(x)]^4}}{L_g V(x)} & \text{si } L_g V(x) \neq 0 \end{cases}$$

Apéndice D

Acerca de la Ecuación 6.27

Sea $S_\infty(\theta)$ la solución estacionaria de:

$$\dot{S}_t(\theta) = -\theta S_t(\theta) - A^T S_t(\theta) - S_t(\theta) A + C^T C$$

donde

$$C = [1 \ 0 \ 0] \quad A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

y $S_t(\theta)$ es una matriz simétrica, positiva definida.

Entonces, $\forall \theta > 0$, la solución estacionaria $S_\infty(\theta)$ satisface

$$(S_\infty(\theta))_{i,j} = (S_\infty(1))_{i,j} \frac{1}{\theta^{i+j-1}} \quad (\text{ecuación 6.27})$$

Demostración

Por simplicidad, realizaremos la demostración para la matriz $S_\infty(\theta)$ de orden 3. Se verá que, análogamente, puede extenderse para matrices de cualquier orden.

Observemos que

$$S_t(\theta) = e^{-\theta t} e^{-A^T t} S_0(\theta) e^{-A t} + \int_0^t e^{-\theta(t-s)} e^{-A^T(t-s)} C^T C e^{-A(t-s)} ds$$

Entonces para t grande $e^{-\theta t} \rightarrow 0$ y se puede “despreciar” el primer término. Luego, para t grande,

$$S_\infty(\theta) = \lim_{t \rightarrow \infty} \int_0^t e^{-\theta(t-s)} e^{-A^T(t-s)} C^T C e^{-A(t-s)} ds$$

Sea el cambio de variables $\rho = t - s$, $d\rho = -ds$.

Como $e^{A^T \rho} \cong I - A^T \rho + (A^T)^2 \frac{\rho^2}{2}$, se tiene:

$$e^{A^T \rho} C^T \cong \begin{bmatrix} 1 \\ -\rho \\ \frac{\rho^2}{2} \end{bmatrix}$$

Con lo cual

$$S_\infty(\theta) = \lim_{t \rightarrow \infty} \int_0^t -e^{-\rho\theta} \begin{bmatrix} 1 \\ -\rho \\ \frac{\rho^2}{2} \end{bmatrix} \begin{bmatrix} 1 & -\rho & \frac{\rho^2}{2} \end{bmatrix} d\rho$$

La matriz $S_\infty(\theta)$ se obtiene calculando la integral anterior para cada $(S_\infty(\theta))_{i,j}$.

Obtenemos de este modo

$$(S_\infty(\theta))_{1,1} = \int_0^\infty e^{-\theta\rho} d\rho = \frac{1}{\theta}$$

$$(S_\infty(\theta))_{1,2} = (S_\infty(\theta))_{2,1} = \int_0^\infty -e^{-\theta\rho} \rho d\rho = \frac{-1}{\theta^2}$$

$$(S_\infty(\theta))_{1,3} = (S_\infty(\theta))_{3,1} = \frac{1}{2} \int_0^\infty e^{-\theta\rho} \rho^2 d\rho = \frac{1}{\theta^3}$$

$$(S_\infty(\theta))_{2,3} = (S_\infty(\theta))_{3,2} = \frac{1}{2} \int_0^\infty -e^{-\theta\rho} \rho^3 d\rho = \frac{-3}{\theta^4}$$

$$(S_\infty(\theta))_{2,2} = \int_0^\infty e^{-\theta\rho} \rho^2 d\rho = \frac{2}{\theta^3}$$

$$(S_\infty(\theta))_{3,3} = \frac{1}{4} \int_0^\infty e^{-\theta\rho} \rho^4 d\rho = \frac{6}{\theta^5}$$

Entonces,

$$S_\infty(\theta) = \begin{bmatrix} \frac{1}{\theta} & \frac{-1}{\theta^2} & \frac{1}{\theta^3} \\ \frac{-1}{\theta^2} & \frac{2}{\theta^3} & \frac{-3}{\theta^4} \\ \frac{1}{\theta^3} & \frac{-3}{\theta^4} & \frac{6}{\theta^5} \end{bmatrix}$$

con

$$S_{\infty}(1) = \begin{bmatrix} 1 & -1 & 1 \\ -1 & 2 & -3 \\ 1 & -3 & 6 \end{bmatrix}$$

y por lo tanto

$$(S_{\infty}(\theta))_{i,j} = (S_{\infty}(1))_{i,j} \frac{1}{\theta^{i+j-1}}$$

como queríamos demostrar.

Apéndice E

Obtención de la cota para la norma de $E(w)$

Para realizar la acotación deseada, graficamos $\|E\|$ en función de z_2 y z_3 para distintos valores de z_1 como parámetro.

E.1 Región de operación elegida para establecer la cota de $E(w)$

Para realizar los gráficos se determinaron los siguientes intervalos en las tres variables de estado:

a) concentración de oxígeno $(-1.5 \text{ mg/l} \leq z_1 \leq 2 \text{ mg/l})$

El intervalo se eligió teniendo en cuenta que la concentración de O_2 es la variable controlada en el sistema, habiéndose fijado su set point en $z_1 = 0.5 \text{ mg/l}$. La cota crece para valores cercanos a -2 mg/l , lo que no es relevante puesto que la concentración de O_2 de equilibrio es $z_1^* = 3.04 \text{ mg/l}$, con lo cual el valor límite tomado indica una concentración muy baja de oxígeno (0.5 mg/l), a los efectos prácticos cercana a cero. En rigor, antes de alcanzar este valor debería producirse un disparo de alarma, para evitar el colapso del sistema por la muerte de gran parte de los microorganismos que producen la reacción biológica. Recordemos que el proceso de lodos activados es aeróbico, y muy sensible a la concentración de O_2 disuelto en el agua.

b) concentración de sustrato $(-2 \text{ mg/l} \leq z_2 \leq 6 \text{ mg/l})$

El valor máximo elegido indicaría que el sistema está proveyendo un

efluente con una carga orgánica que es el doble de la esperada. Con esto se estarían incumpliendo largamente las especificaciones de calidad requeridas, por lo cual es deseable que antes de alcanzar esta condición el sistema de detección de fallas haya previsto una señal de alarma indicando el comportamiento anormal del proceso. En cuanto al mínimo, se tuvo en cuenta que en un proceso de lodos activados tradicional se logra una reducción de la DBO de hasta un 5 y un 15% de su valor inicial en la alimentación, es decir un rendimiento en el consumo de la DBO soluble del 85 al 95 %. El valor mínimo elegido en este caso corresponde a reducir la carga orgánica hasta aproximadamente un 1% de su valor en la alimentación. Por otra parte, se realizaron cálculos para llevar prácticamente a cero la concentración de sustrato en el biorreactor, observándose que esto no implicaría un cambio significativo en la cota elegida para E .

c) concentración de biomasa ($-300 \text{ mg/l} \leq z_3 \leq 300 \text{ mg/l}$)

En este caso, el intervalo elegido está en estrecha correlación con los anteriores. Si el proceso se encuentra funcionando en condiciones “normales”, no es esperable que la concentración de microorganismos esté por fuera de este intervalo. Situaciones especiales, tales como el agotamiento del biorreactor (disminución de la biomasa a cero y concentración de sustrato en la salida igual a la de la alimentación) deberían ser detectadas con un disparo de alarma por el sistema de DEF.

Recuérdese que las variables z_1, z_2, z_3 representan concentraciones relativas a las de equilibrio.

E.2 Figuras para la determinación de la cota

Los resultados se muestran en las Figuras E.1 a E.9. Observamos la forma de $\| E \|$ para valores de z_2, z_3 en los rangos indicados y z_1 como parámetro. Se puede ver que si admitimos que (z_1, z_2, z_3) pertenecen al compacto $[-1.5, 2] \times [-2, 6] \times [-300, 300]$ entonces $\| E \| \leq 550$ (Véase la figura E.8).

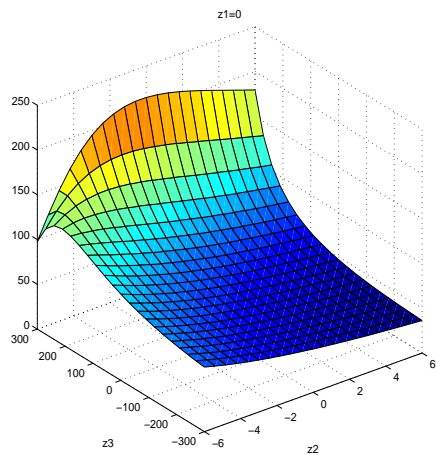


Figura E.1: Cota de $\|E\|$ cuando el valor de z_1 es el de equilibrio

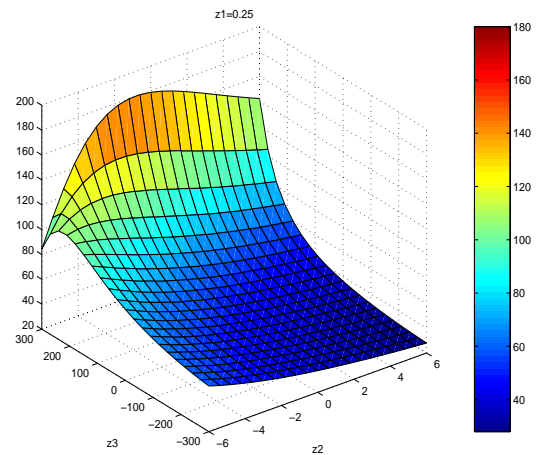


Figura E.2: Cota de $\|E\|$ cuando $z_1 = 0.25$

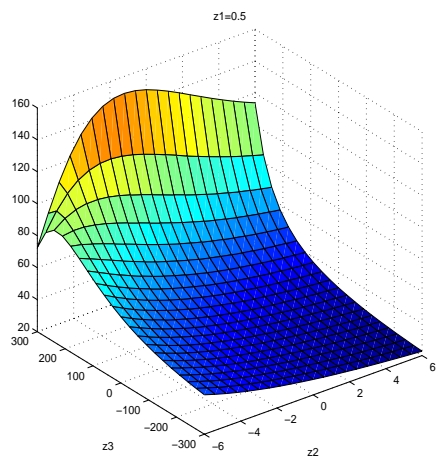


Figura E.3: Cota de $\|E\|$ cuando $z_1 = 0.5$

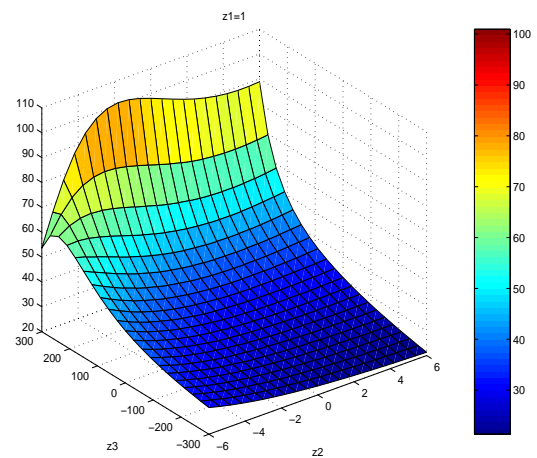


Figura E.4: Cota de $\|E\|$ cuando $z_1 = 1$

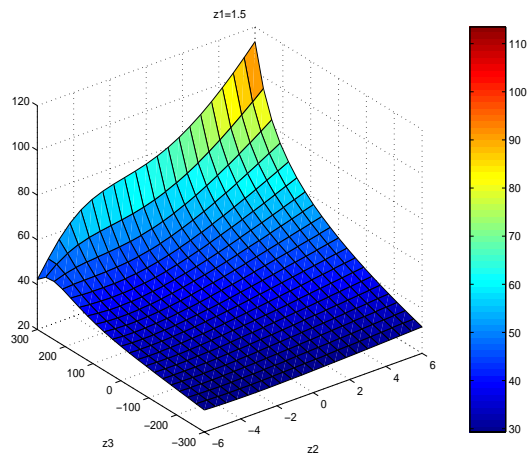


Figura E.5: Cota de $\| E \|$ cuando $z_1 = 1.5$

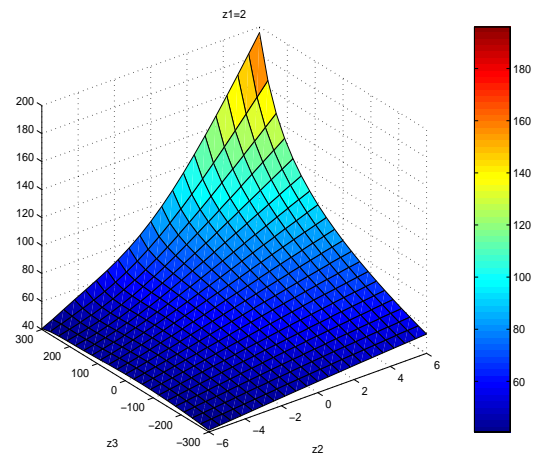


Figura E.6: Cota de $\| E \|$ cuando $z_1 = 2$

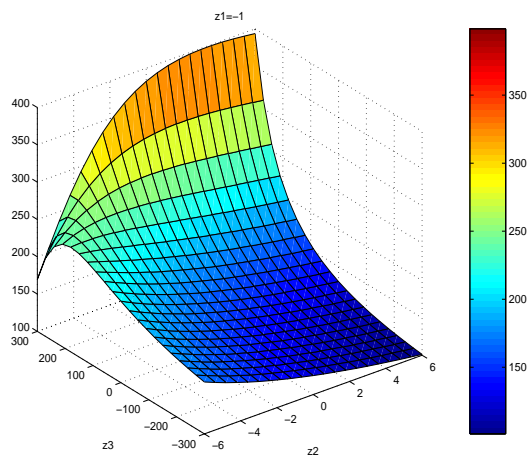


Figura E.7: Cota de $\| E \|$ cuando $z_1 = -1$

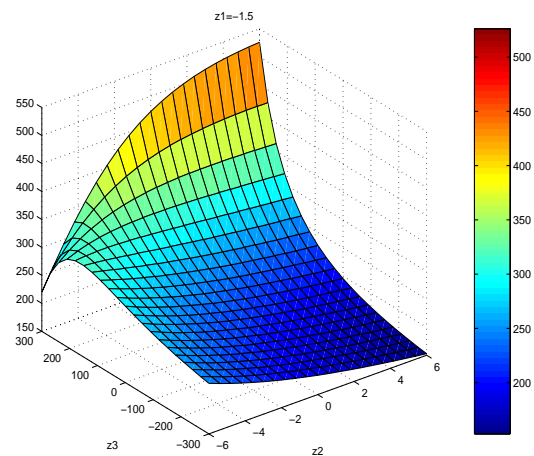


Figura E.8: Cota de $\| E \|$ cuando $z_1 = -1.5$

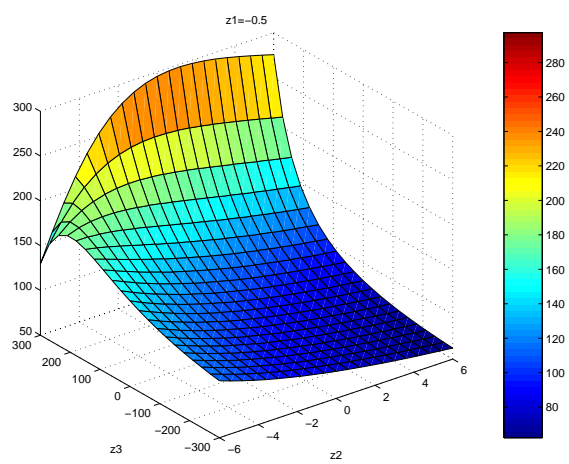


Figura E.9: Cota de $\| E \|$ cuando $z_1 = -0.5$

Bibliografía

- [1] Alcorta García E. y Frank P., “Deterministic nonlinear observer-based approaches to fault detection: a survey,” *Control Eng. Practice*, Vol. 5, Nro. 5, pp. 663-670, 1997.
- [2] Aubrun C., Harmand J., Garnier O., Steyer J., “Fault detection filter for an anerobic digestion process,” *Bioprocess Engineering*, Vol. 22, pp. 413-420, 2000.
- [3] Basseville M. Nikiforov I.V., *Detection of abrupt changes: Theory and Application*, Information and System Science Series, Prentice Hall, New York, 1993.
- [4] Basseville M., “Detecting changes in signals and systems, a survey,” *Automatica*, Vol. 24, Nro. 3, pp. 309-326, 1988.
- [5] Basseville M. y Benveniste A., *Detection of abrupt changes in signals and dynamical systems*, Lecture Notes in Control and Information Sciences, N°77, Springer- Verlag, 1986.
- [6] Beard R. V., *Failure accomodation in Linear system through self reorganization*, PhD Thesis, Massachusetts Institute of Technology, Mass., USA, 1971.
- [7] Bertola N., Palladino L., Bevilacqua A., Zaritzky N., “Optimisation of the design parameters in an activated sludge system for the wastewater treatment of a potato processing plant,” *Journal of Food Engineering*, Vol. 40, pp. 27-33, 1999.
- [8] Biagiola S., Solsona J., Milocco R., “Two different approaches for biomass estimation in batch bioprocesses,” *Anales de la VIII Reunión de Trabajo en Procesamiento de la Información y Control*, Mar del Plata, Argentina, 1999, Vol. 3, pp. 79-87.

- [9] Boskovic J., "Observer-based adaptive control of a class of biorreactor processes," *Proc. of the 34th. Conf. on Decision and Control*, 1995, pp. 1171–1176.
- [10] Carlsson, B. and R. Milocco, "A simple strategy for controlling the nitrate concentration in an activated sludge process using external carbon flow rate," *Anales de la VIII Reunión de Trabajo en Procesamiento de la Información y Control*, Mar del Plata, Argentina, 1999, Vol. 9, pp. 21-30.
- [11] Carlsson B., Hasselblad S., Plaza E., Mårtensson S., Lindberg C., "Design and operation of a pilot-scale activated sludge plant," *Vatten*, Vol. 53, pp. 27-32, 1997.
- [12] Clark R. N., "A simplified instrument failure detection scheme," *IEEE Trans. AES*, Vol. AED-14, pp. 558-563, 1978.
- [13] Chen J. y Patton R.J., *Robust model based fault detection for dynamic systems*, Kluwer Academic Publishers, 1999.
- [14] Chen J., Patton R.J., and Zhang H.Y., "Design of unknown input observers and robust fault-detection filters," *Int. J. Contr.*, Vol. 63, Nro. 1, pp. 85–105, 1996.
- [15] Chow E. and Willsky A., "Issues in the development of a general algorithm for reliable failure detection," *Proc. of the 19th. IEEE Conf. on Dec. and Control*, Albuquerque, 1980.
- [16] D'Attellis C.E., *Introducción a los sistemas no lineales de control y sus aplicaciones*, AADECA, 1992.
- [17] De Persis, C. and Isidori, A. "A geometric approach to nonlinear Fault detection and Isolation," *IEEE Trans. on Autom. Control*, Vol. 46, no. 6, pp. 853-865, Jun. 2001 .
- [18] Derco J., Kovacs A., Králik M., Shansab S., "Modelling of an intermittently aerated completely mixed activated sludge reactor," *Mathematical and Computer Modelling of Dynamical Systems*, Vol. 5, Nro. 4, pp. 313-336, 1999.
- [19] Ding X., Guo L., Frank P.M., "A frequency domain approach to fault detection of uncertain dynamic systems," *Proc. of the 32th. IEEE Conf. on Dec. and Control*, Texas, 1993, pp. 1722–1727.

- [20] Ding X., y Frank P. M., "Fault detection via factorization approach," *Syst. Cont. Lett.*, Vol. 14, Nro. 5, pp. 431–436, 1990.
- [21] Dochain D., "Design of adaptive controllers for non-linear stirred tank bioreactors: extension to the MIMO situation," *J. Proc. Cont.*, Vol. 1, pp. 41–48, 1991.
- [22] Farza M., Busawon K. and Hammouri H., "Simple non linear observer for on-line estimation of kinetic rates in bioreactors," *Automatica*, Vol. 34, Nro. 3, pp. 301–318, 1998.
- [23] Frank P. y Ding, X., "Survey of robust residual generation and evaluation methods in observer-based fault detection systems," *J. Proc. Cont.*, Vol. 7, Nro. 6, pp. 403–424, 1997.
- [24] Frank P.M., "Analytical and qualitative model-based fault diagnosis- a survey and some new results," *European J. of Contr.*, Vol. 2, Nro. 1, pp. 6–28, 1996.
- [25] Frank P., "Fault diagnosis in dynamic systems using analytical and knowledge-based redundancy—A survey and some new results," *Automatica*, Vol. 26, Nro. 3, pp. 459–474, 1990.
- [26] García R. Troparevsky M.I. y Mancilla Aguilar J.L., "An observer for Nonlinear noisy systems," *Latin American Applied Research*, Vol. 30, pp. 87–92, 2000.
- [27] Gauthier J.P., Bornard P., "Observability for any $u(t)$ of a class of nonlinear systems," *IEEE Trans. Aut. Contr.*, Vol. AC-26, Nro. 4, pp. 922–926, 1981.
- [28] Gauthier J.P., Hammouri H. y Othman, S., "A simple observer for nonlinear systems. Applications to bioreactors," *IEEE Trans. Aut. Cont.*, Vol. 37, Nro. 6, pp. 875–880, 1992.
- [29] Ge W. and Fang C., "Detection of faulty components via robust observation," *Int. J. Contr.*, Vol. 47, Nro. 2, pp. 581–599, 1988.
- [30] Han Zhengzhi, Chen Guanrong, "Dynamic right coprime factorization and observer design for nonlinear systems," *Latin American Applied Research*, Vol. 32, pp. 327–336, 2002.
- [31] Henze M., P. Harremoës, J. Jansen, E. Arvin, *Wastewater Treatment, Biological and Chemical Processes*, 2nd. ed., Springer, 1997.

- [32] Isermann R., "Process fault detection based on modelling and estimation methods: A survey," *Automatica*, Vol. 20, Nro. 4, pp. 387-404, 1984.
- [33] Isidori A., *Nonlinear control systems*, 3rd. ed., Springer-Verlag, 1996.
- [34] The IWA Task Group on Mathematical Modelling for design and operation of biological wastewater, *Activated sludge models: ASM1, ASM2, ASM2d, ASM3*, IWA Publishing, 2000.
- [35] Jones H.L., *Failure Detection in linear systems*, PhD thesis, Massachusetts Institute of Technology, Mass., USA , 1973.
- [36] Julien S., Lessard P., Barbary J.P., "A reduced order model for control of a single reactor activated sludge process," *Mathematical and Computer Modelling of Dynamical Systems*, Vol. 5, Nro. 4, pp. 337-350, 1999.
- [37] Lefschetz S., *Differential Equations: Geometric Theory*, Dover Pub., New York, 1977.
- [38] Li-Cheng Shen y Pau-Lo Hsu, "Robust design of fault isolation observers," *Automatica*, Vol. 34, Nro. 14, pp. 1424-1429, 1998.
- [39] Lindberg C., Carlsson B., "Adaptive control of external carbon flow rate in an activated sludge process," *Wat. Sci. Tech.*, Vol. 34, Nro. 3-4, pp. 173-180, 1996.
- [40] Lindberg C., Carlsson B., "Estimation of the respiration rate and oxygen transfer function utilizing a slow DO sensor," *Wat. Sci. Tech.*, Vol. 33, Nro. 1, pp. 325-333, 1996.
- [41] Marino R. and Tomei P., *Nonlinear control design*, Prentice-Hall, 1995.
- [42] Mironovski L.A., "Functional diagnosis of dynamic system- a survey," *Autom. Remote Control*, Vol. 41, pp. 1122-1143, 1980.
- [43] Mouyon P., "Tools for nonlinear observer design," *IEEE Int. Symp. on Diagnostic for Electrical Machines Power Electronics and Drives SDEMPED'97*, Francia, Sep. 1997.
- [44] Olsson G. y Newell B., *Wastewater Treatment Systems -Modelling, Diagnosis and Control*, IWA Publishing, 1999.

- [45] Olsson G., Aspegren H., Nielsen M., "Operation and control of wastewater treatment - a scandinavian perspective over 20 years", *Water Science and Technology*, Vol. 37, no. 12, pp. 1-13, 1998.
- [46] Patton R.J., Frank P.M., Clark R.N., (eds.), *Fault diagnosis in dynamic systems, theory and application*, Control Engineering Series, Prentice Hall, New York, 1989.
- [47] Patton R.J., Chen J. " A review of parity space approaches to fault diagnosis", *IFAC/IMACS Sympo: SAFEPROCESS'91*, Vol. 1, pp, 239-255, Baden-Baden, 1991.
- [48] Patton R., Uppal F., y Lopez-Toribio C., "Soft computing approaches to fault diagnosis for dynamic systems: a survey", *Proc. IFAC Symposium SAFEPROCESS 2000*, Budapest, June 2000, pp. 298-311.
- [49] Ramalho R.S., *Tratamiento de aguas residuales*, Ed. Reverté, 1993.
- [50] Schreier G., Ragot J., Patton R. y Frank P., "Observer design for a class of nonlinear systems," *Proc. of the IFAC Symposium SAFEPROCESS'97*, Univ. of Hull, UK, pp. 483-488, 1997.
- [51] Schuster E.J., Solsona J. and Muravchik C., "Detección de fallas de actuador empleando un observador con dinámica lineal del error," *Anales del XVI Congreso Argentino de Control Automático*, Buenos Aires, Argentina, 1998, pp. 615-619.
- [52] Seliger R. y Frank P., "Fault diagnosis by disturbance decoupled nonlinear observers," *Proc. of the 30th. Conf. on Dec. and Cont. CDC'91*, Brighton, UK, 1991, pp. 2248-2253.
- [53] Solsona J., *Técnicas de control no lineal en sistemas electrónicos de potencia*, Tesis doctoral, Facultad de Ingeniería, UNLP, 1995.
- [54] Suárez M. y Muravchik C., "Observador para la detección de fallas en una planta de tratamiento de aguas residuales. Atenuación de perturbaciones," *Anales de la IX Reunión en Procesamiento de la Información y Control*, Santa Fe, Argentina, 2001.
- [55] Suárez M. y Muravchik C., "Observadores para la detección de fallas en sistemas de tratamiento de aguas residuales," *Anales del XVII Congreso Argentino de Control Automático AADECA 2000*, Buenos Aires, Argentina, Sep. 2000, pp. 363-368.

- [56] Tenno R., Uronen P., "Two sludge nitrification process control, state and parameter estimation," *Proc. of the IFAC 13th. Triennial World Congress*, San Francisco, USA, 1996.
- [57] Troch I. (Editor), Special Issue in Modelling for Control of Wastewater Treatment Plants, *Mathematical and computer modelling of dynamical systems*, Vol. 5, Nro. 4, 1999.
- [58] Vega Cruz P. y De la Fuente M., "Redes neuronales aplicadas a la detección y diagnóstico de fallos," *Aplicaciones de las redes neuronales en supervisión, diagnosis y control de procesos*, CYTED/CONICIT, Ed. Equinoccio, Venezuela, 1999, pp. 69-90.
- [59] Vemuri A. y Polycarpou M., "Robust nonlinear fault diagnosis in input-output systems," *Int. J. Control*, Vol. 68, Nro. 2, pp. 343-360, 1997.
- [60] Walcott B.L. and S.H. Zak, "State observation of nonlinear uncertain dynamical systems," *IEEE Trans. on Autom. Contr.*, Vol. 32, Nro. 2, pp. 166-169, 1987.
- [61] Wang Y., Davison E., Dorato P., "Observing the states of systems with unmesurable disturbance," *IEEE Trans. Automat. Contr.*, Vol. AC-20, pp. 716-717, 1975.
- [62] Wei Ge y Chong-Zhi Fang, "Detection of faulty components via robust observation," *Int. J. Control*, Vol. 47, Nro. 2, pp. 581-599, 1988.
- [63] Willems J.C., "The least squares stationary optimal control and the algebraic Riccati equation," *IEEE Trans. Autom. Control*, Vol. 16, pp. 621-634, 1971.
- [64] Willsky, A.S., "A survey of design methods for failure detection in dynamic systems," *Automatica*, Vol. 12, Nro. 6, pp. 601-611, 1976.
- [65] Wünnenberg J., *Observer-based fault detection in dynamic systems*, PhD Thesis, Univ. of Duisburg, Germany, 1990.
- [66] Yaz E. y Azemi A., "Actuator fault detection and isolation in nonlinear systems using LMI's and LME's," *Proc. of the American Control Conf.*, Philadelphia, USA, 1998, pp. 1590-1594.
- [67] Zhao H., Mc Avoy T., "Modeling of activated sludge wastewater treatment processes using integrated neural networks and a first principle model," *Proc. IFAC 13th Triennial World Congress*, San Francisco, USA, 1996, pp. 455-460.

- [68] Zolghadri A., Henry D. y Monsion M., “Design of nonlinear observers for fault diagnosis. A case study,” *Cont. Eng. Practice*, Vol. 4, Nro. 11, pp. 1535–1544, 1996.

Correcciones sugeridas

1. pág. 27 y pág. 31: Se agregaron las referencias sugeridas.
2. pág. 37: Se corrigió la ecuación (3.34) y se agregó la expresión de $r(s)$.
3. pág. 50: Comentario a la ecuación (3.74)
4. pág. 65: Se agregó una nota con la definición de las derivadas de Lie.
5. pág. 85: Se corrigió la tabla (5.1).
6. pág. 104 y 109: Se modificó la notación para mantener consistencia.
7. pág. 107: (OEV)
8. pág. 117: En las figuras, se grafican las concentraciones referidas a las de equilibrio (ver figuras), por lo cual pueden aparecer valores negativos de las variables.

Correcciones sugeridas

1. pág. 27 y pág. 31: Se agregaron las referencias sugeridas.
2. pág. 37: Se corrigió la ecuación (3.34) y se agregó la expresión de $r(s)$.
3. pág. 50: Comentario a la ecuación (3.74)
4. pág. 65: Se agregó una nota con la definición de las derivadas de Lie.
5. pág. 85: Se corrigió la tabla (5.1).
6. pág. 104 y 109: Se modificó la notación para mantener consistencia.
7. pág. 107: (OEV)
8. pág. 117: En las figuras, se grafican las concentraciones referidas a las de equilibrio (ver figuras), por lo cual pueden aparecer valores negativos de las variables.